VOLUME 107 PART B NUMBER 32 MARCH 1960

J. T. LUDWIG
APR 23 1960
SCIENTIFIC ADVISORY GROUP



The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

ELECTRONIC AND COMMUNICATION ENGINEERING
(INCLUDING RADIO ENGINEERING)

SAVOY PLACE · LONDON W.C.2

Price Ten Shillings and Sixpence

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1959–1960

President

SIR WILLIS JACKSON, D.Sc., F.R.S.

Past-Presidents

W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROF. E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc. (Eng.), LL.D.
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROF. E. B. MOULLIN, M.A., Sc.D., LL.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COL. B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.), F.C.G.I.
SIR JOSIAH ECCLES, C.B.E., D.Sc.
THE RT. HON. THE LORD NELSON OF STAFFORD
SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).
S. E. GOODALL, M.Sc.(Eng.), F.Q.M.C.

Vice-Presidents

O. W. HUMPHREYS, C.B.E., B.Sc. G. S. C. LUCAS, O.B.E., F.C.G.I. SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc. C. T. MELLING, C.B.E., M.Sc.Tech.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).

Honorary Treasurer

E. LEETE.

Ordinary Members of Council

PROF. H. E. M. BARLOW, Ph.D., B.Sc. (Eng.).
C. O. BOYSE, B.Sc.(Eng.).
PROF. M. W. HUMPHREY DAVIES, M.Sc.
SIR JOHN DEAN, B.Sc.
L. DRUCQUER.
J. M. FERGUSON, B.Sc.(Eng.).
D. C. FLACK, B.Sc.(Eng.), Ph.D.
J. S. FORREST, D.Sc., M.A.
R. J. HALSEY, C.M.G., B.Sc.(Eng.), F.C.G.I.
J. B. HIGHAM, Ph.D., B.Sc.
R. A. HORE, M.A., B.Sc.
F. C. MCLEAN, C.B.E., B.Sc.
B. L. METCALF, B.Sc.(Eng.).
J. R. MORTLOCK, Ph.D., B.Sc.(Eng.).
THE HON. H. G. NELSON, M.A.
R. H. PHILLIPS, T.D.
H. V. PUGH.
J. R. RYLANDS, M.Sc., J.P.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.).
C. E. STRONG, O.B.E., B.A., B.A.I.
D. H. TOMPSETT, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Sections

Electronics and Communications:
M. J. L. PULLING, C.B.E., M.A.
*G. MILLINGTON, M.A., B.Sc.

Measurement and Control:
PROF. A. TUSTIN, M.Sc.
*J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.

Supply:
J. R. MORTLOCK, Ph.D., B.Sc.(Eng.).
*D. P. SAYERS, B.Sc.

Utilization:
T. E. HOUGHTON, M.Ent.
*R. A. MARRYAT, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
D. H. PARRY, B.Sc.
*D. E. LAMBERT, B.Sc.(Eng.).

Mersey and North Wales Centre:
T. A. P. COLLEDGE, B.Sc.(Eng.).
*J. COLLINS.

North-Eastern Centre:
H. WATSON-JONES, M.Eng.
*A. T. CRAWFORD, B.Sc.

North Midland Centre:
PROF. G. W. CARTER, M.A.
*J. D. NICHOLSON, B.Sc.

North-Western Centre:
F. H. HUTCHINSON, M.Eng.
*PROF. F. C. WILLIAMS, O.B.E., D.Sc., D.Phil., F.R.S.

Northern Ireland Centre:
T. S. WYLIE.
*D. S. MCILHAGGER, Ph.D., M.Sc.

Scottish Centre:
J. A. AKED, M.B.E.
*R. J. RENNIE, B.Sc.

South Midland Centre:
G. F. PEIRSON.
*L. L. TOLLEY, B.Sc.(Eng.).

Southern Centre:
W. D. MALLINSON, B.Sc.(Eng.).
*G. BISHOP, B.Sc.

Western Centre:
H. JACKSON, B.Sc.(Eng.).
*R. W. STEEL.

* Past-Chairman

ELECTRONICS AND COMMUNICATIONS SECTION COMMITTEE 1959–1960

Chairman

M. J. L. PULLING, C.B.E., M.A.

Vice-Chairmen

R. J. HALSEY, C.M.G., B.Sc.(Eng.), F.C.G.I. T. B. D. TERRONI, B.Sc. J. A. RATCLIFFE, O.B.E., M.A., F.R.S.

Past-Chairmen

G. MILLINGTON, M.A., B.Sc. J. S. MCPETRIE, Ph.D., D.Sc.

Ordinary Members of Committee

D. A. BARRON, M.Sc.
P. A. T. BEVAN, C.B.E., B.Sc.
J. BROWN, M.A., Ph.D.
PROFESSOR A. L. CULLEN, Ph.D., B.Sc.(Eng.).
L. I. FARREN, M.B.E.
G. G. GOURIET.
COMDR. C. G. MAYER, O.B.E., U.S.N.R.
J. MOIR
L. J. I. NICKELS, B.Sc.(Eng.).
B. G. PRESSEY, M.Sc.(Eng.). Ph.D.
N. C. ROLFE, B.Sc.(Eng.).
T. R. SCOTT, D.F.C., B.Sc.
C. WILLIAMS, B.Sc.(Eng.).
W. E. WILLSHAW, M.B.E., M.Sc.Tech.
A. J. YOUNG, B.Sc.(Eng.).

And

The President (*ex officio*)
The Chairman of the Papers Committee.
PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.) (representing the Council).
E. H. COOKE-YARBOROUGH, M.A. (Co-opted Member).
G. H. HICKLING, B.Sc. (representing the North-Eastern Measurement and Electronics Group).
K. F. SANDER, M.A., Ph.D., B.Sc. (representing the Cambridge Electronics and Communications Group).
J. STEWART, M.A., B.Sc. (representing the Scottish Electronics and Measurement Group).
R. FEINBERG, Dr.Ing., M.Sc (representing the North-Western Electronics and Communications Group).
R. E. YOUNG, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).

The following nominees:
Royal Navy: CAPTAIN W. D. F. B. MUSPRATT, R.N.
Army: COL. R. G. MILLER, M.A.
Royal Air Force: GROUP CAPTAIN D. W. ROWSON, B.Sc.(Eng.), R.A.F.

MEASUREMENT AND CONTROL SECTION COMMITTEE 1959–1960

Chairman

PROFESSOR A. TUSTIN, M.Sc.

Vice-Chairmen

C. G. GARTON; W. S. ELLIOTT, M.A.

Past-Chairmen

J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.; H. S. PETCH, B.Sc.(Eng.).

Ordinary Members of Committee

E. W. CONNON, B.Sc.(Eng.), M.Eng.
A. C. LYNCH, M.A., B.Sc.
A. J. MADDOCK, D.Sc.
R. E. MARTIN.
A. NEMET, Dr.Sc.Techn.
S. N. POCOCK.
W. RENWICK, M.A., B.Sc.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.).
G. F. TAGG, Ph.D., B.Sc.
R. D. TROTTER, B.Sc.(Eng.).
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
F. C. WIDDIS, B.Sc.(Eng.).

And

The President (*ex officio*).
The Chairman of the Papers Committee.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.) (representing the Council).
C. C. BAXENDALE (representing the North-Eastern Measurement and Electronics Group).
E. J. R. HARDY, B.Sc.(Eng.) (representing the North-Western Measurement and Control Group).
D. L. A. BARBER, B.Sc.(Eng.) (nominated by the National Physical Laboratory).
H. M. GALE, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).
W. H. P. LESLIE, B.Sc. (representing the Scottish Electronics and Measurement Group).

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

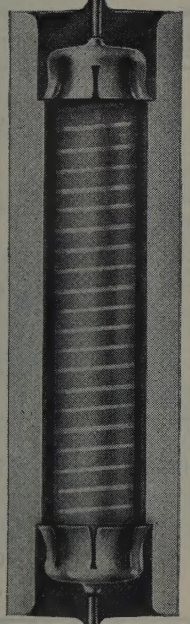
Principal Assistant Secretary

F. C. HARRIS

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E

Fully Protected HIGH STABILITY Resistors



The Inside Story

There is, of course, a lot of "know-how" in the manufacture of a high stability resistor element to the exacting characteristics demanded of a high stability resistor, but, having made the element, two problems remain. How to protect the smoke thin resistance film from damage in transit, in handling, and in assembly, and how to isolate the element from contact with paints, lacquers, and other finishes, all of which have a tendency to pull the film under extremes of temperature and humidity.

In the Erie high stability resistor these two problems have been solved very simply and with complete effectiveness by the encapsulation of the element in the ceramic insulating tube, cement sealed at the ends, proven on billions of Erie solid carbon resistors and ceramic dielectric capacitors in use throughout the world. This tube obviously affords complete protection from all manner of physical damage, and, as can be seen from the illustration, the counter bore at either end and the close affinity between the counter bore and the caps of the element ensures that the element is supported by the caps clear of the inner bore of the tube, and there is thus no contact whatsoever with any material that might prove harmful.

Only Erie high stability resistors are protected in this way, and that is why they are found in all equipment where robustness and reliability under all conditions must be allied to first-class performance.

ERIE* Resistor Ltd

★ Registered Trade Marks

I, HEDDON STREET, LONDON, W.I., Tel. REGent 6432

Factories: Great Yarmouth and Tunbridge Wells, England; Trenton, Canada; Erie, Pa., Holly Springs, Miss. and Hawthorne, Cal., U.S.A.

Semiconductors

COMPUTER TRANSISTORS



The Semiconductors range of Computer Transistors, designed and tested to the special requirements of computer engineers, is the key to a new order of computer speed and reliability. Overall reliability is further increased by making possible a substantial reduction in the number of associated components.

The two types of Silicon Alloy Transistor now in full-scale production make it possible to extend this high-speed computer performance into **ambient** temperatures well above 100°C. **Samples are available now.**

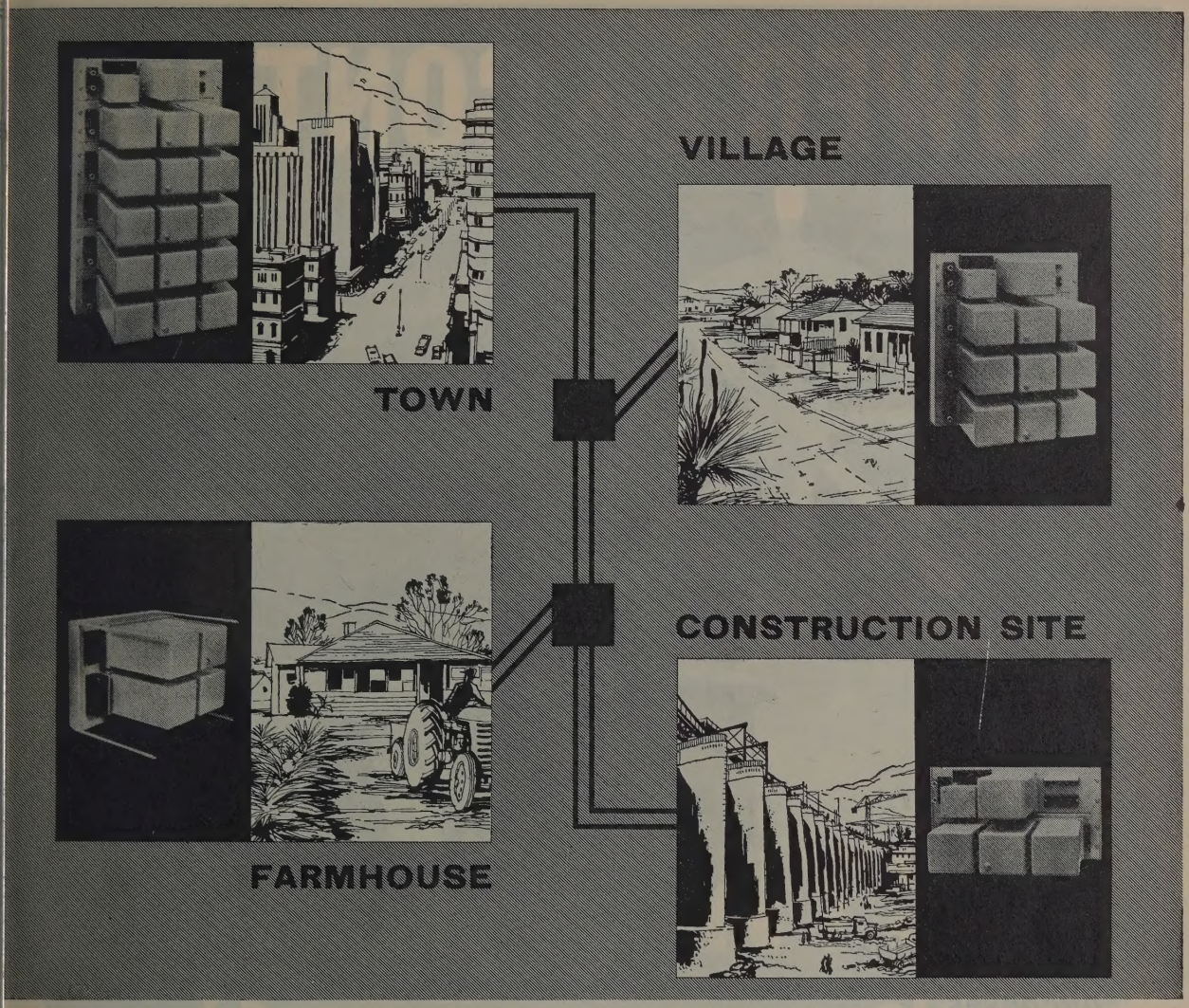
	TYPE	DESCRIPTION	RISE TIME millimicroseconds	Vc max	Ic max
HIGH-SPEED LOW-LEVEL SWITCHING GERMANIUM	SB 344 SB 345	General purpose transistors for conventional logic circuits.	50	5v	5mA
	SB 240	Designed for directly coupled circuits. Controlled input, saturation and hole storage characteristics.	30	6v	15mA
	MA 393	High gain transistor for high-speed driving of parallel circuits.	30	6v	50mA
	2N 501	Ultra-high speed transistor with controlled input and saturation characteristics.	10	12v	50mA
HIGH-SPEED LOW-LEVEL SWITCHING SILICON	SA 495	General purpose 10Mc/s transistor for conventional logic circuits.	100	25v	50mA
	SA 496	15Mc/s transistor for directly coupled circuits. Saturation resistance typically 10 ohms. Controlled input and hole storage characteristics.	80	10v	50mA
CORE DRIVING GERMANIUM	2 N 597 2 N 598 2 N 599	min $f\alpha$ 3Mc/s min $f\alpha$ 5Mc/s min $f\alpha$ 12Mc/s 250 mW high frequency alloy transistors with high gain and low saturation resistance	400 * 250 * 100 *	45v 30v 20v	400mA 400mA 400mA
	2 N 600 2 N 601	min $f\alpha$ 5Mc/s min $f\alpha$ 12Mc/s 750 mW versions of 2 N 598 and 2 N 599.	250 * 100 *	30v 20v	400mA 400mA

* *rise time to 400mA*

Full technical details and applications assistance available on request.

Semiconductors Limited

CHENEY MANOR
SWINDON · WILTS
TELEPHONE: SWINDON 6421



Transistorised rural carrier telephone system-TRC7

in service in many parts of the world

- Fully transistorised
- Simple installation
- Low power consumption
- Up to seven extra speech channels
- Ideal for temporary extra lines
- Fully tropicalised
- Pole mounted drop-off filters for greater flexibility

A PRODUCT OF

MULLARD EQUIPMENT LIMITED

A COMPANY OF THE MULLARD GROUP

Write for descriptive literature

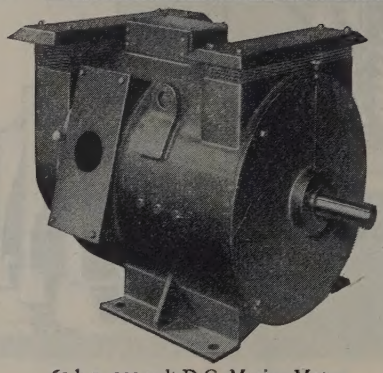
Mullard House · Torrington Place
London · W.C.1 · Tel: LANgham 6633

ME626c

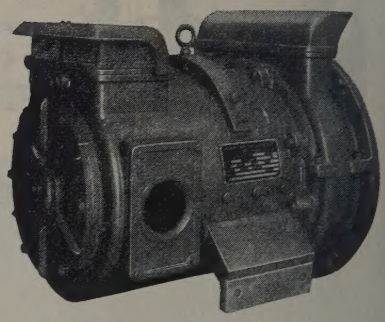
POWER AND CONTROL



E.D.C.C. make Motors, Generators and Starters for all marine purposes, in standard and special designs.

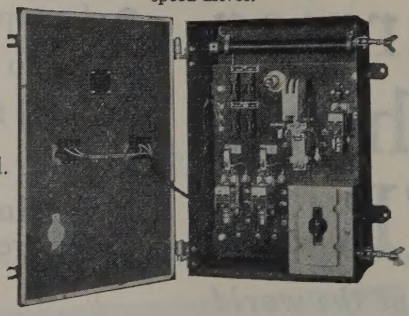


60 h.p. 220 volt D.C. Marine Motor, drip-proof enclosure.

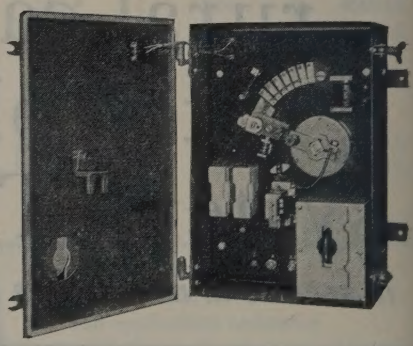


35 kW. 220 volt D.C. Marine Generator, flange mounted for coupling to diesel engine.

Contactor Starter for D.C. motors available for constant and variable speed drives.



Hand operated Faceplate Starter for D.C. motors, available for constant and variable speed drives.



Illustrated here is equipment especially designed for marine duties and backed by long and successful experience in this field.

This equipment complies fully with all relevant B.S. specifications, the requirements of Lloyds and British M.O.T.

PLEASE SEND FOR FURTHER DETAILS

ELECTRO-DYNAMIC

CONSTRUCTION COMPANY LIMITED

ST MARY CRAY ORPINGTON KENT TELEPHONE: ORPINGTON 27551 TELEGRAMS: ELEDAMIC ST MARY CRAY

Control Gear Division: BRIDGWATER SOMERSET Tel: BRIDGWATER 2882 Glasgow Office: 40 HOULDSWORTH ST C3 Tel: CENTRAL 2620

P.3969



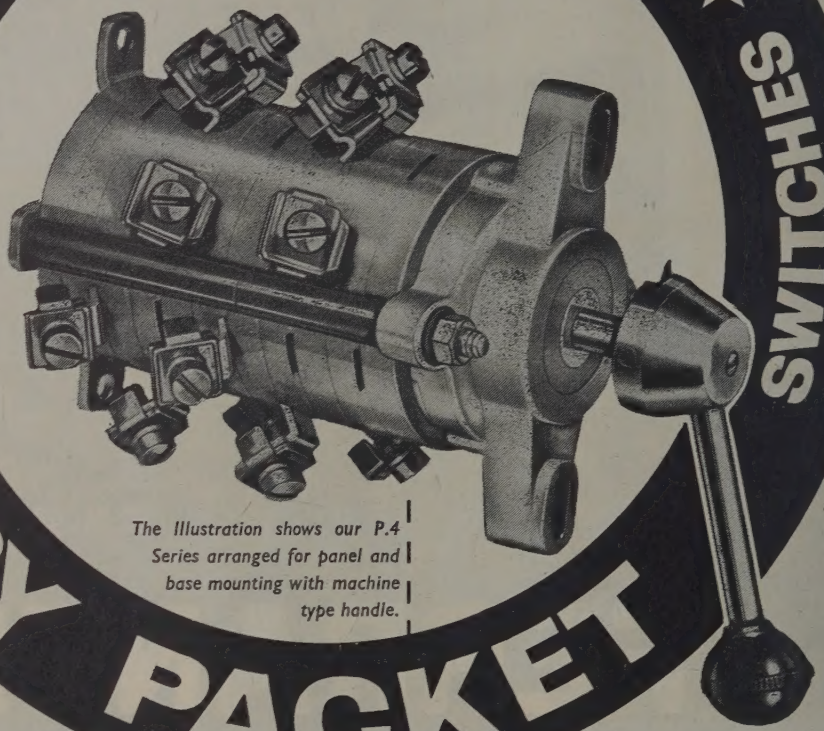
Leading Radar Manufacturers use Ferranti T.R. Cells in Airborne, Ground and Marine Radar Systems. A comprehensive range of T.R. Cells is available covering frequencies from 1000 Mc/s. to 35,000 Mc/s.
Write for further information to:

FERRANTI LTD · KINGS CROSS ROAD · DUNDEE Tel: DUNDEE 87141

Ferranti

DS/T64

★ MULTI-CIRCUIT ★ ROTARY PACKET SWITCHES



The Illustration shows our P.4 Series arranged for panel and base mounting with machine type handle.

ARROW Rotary Packet switches have, for many years, enjoyed an enviable reputation for excellence of design and fine performance.

Considerable saving in cost can be effected by the simultaneous switching of various circuits using ARROW Rotary switches.

QUICK DELIVERY

Write TODAY for a copy of LIST R.P.2. describing our P.4 Series illustrated here.

P.4 SERIES

Maximum Non-Reactive Ratings

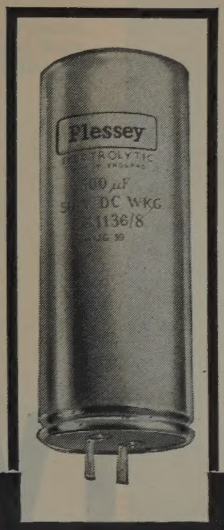
30/35 Amps 250v AC/DC
20/30 Amps 440v AC/DC
15/25 Amps 600v AC/DC

ARROW

ARROW ELECTRIC SWITCHES LTD

HANGER LANE · LONDON W5

A/31



Plessey

HYPERLYTICS* give

ten years service

Long shelf life and a working life of 10 years are achieved in the new type of electrolytic capacitor developed by Plessey.

The exacting specifications demanded for specialised electronic equipment have been met by using a high purity etched foil material, and combining the direct Plessey process control with unique manufacturing methods.

With low leakage and high insulation characteristics, Hyperlytics can be supplied in a wide range of values between 0.5 and 12,000 μF. Two types are available, the 1000 series with a temperature range of —30°C to +85°C, and the 11,000 series with a range of —20°C to +50°C.

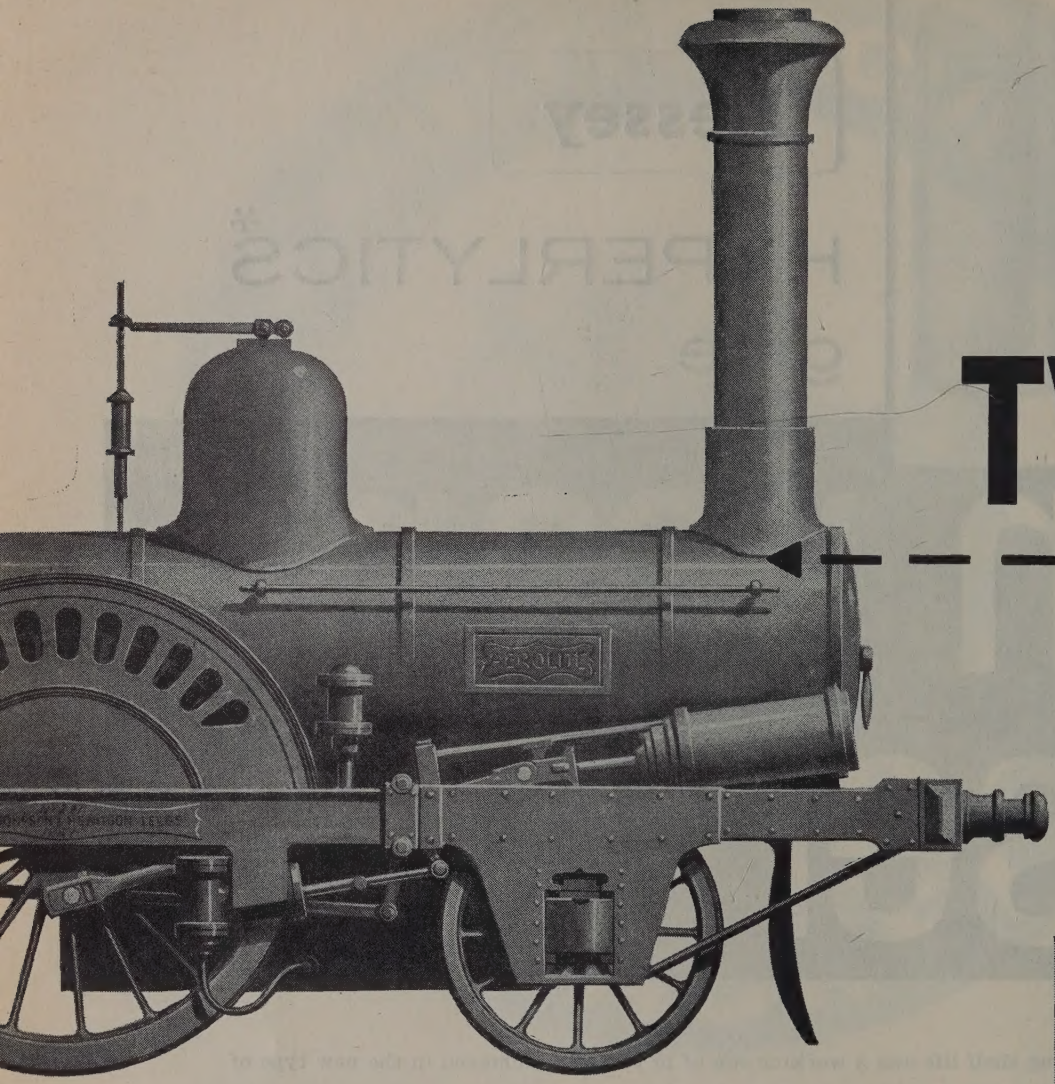
Capacitance tolerance for both types is normally —20% to +50% but closer limits can be achieved to meet special requirements.

Write for further details of these outstanding new components.

* *High Performance Electrolytic Capacitors*

THE PLESSEY COMPANY LIMITED · CAPACITORS & RESISTORS DIVISION
KEMBREY STREET · SWINDON · WILTS · TELEPHONE: SWINDON 6211
Overseas Sales Organisation: Plessey International Limited · Ilford · Essex · Telephone: Ilford 3040

CR1/A



TWO 108

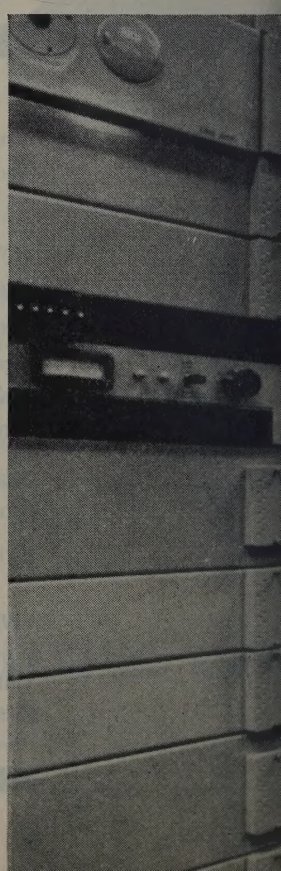
On 15th May 1852 the formal opening took place of the extension of the Leeds Northern Line to Stockton, establishing the first railway link between Leeds and Middlesbrough. At that time only the most elementary line telegraph communications existed.

In 1960, 108 years later, a further link has been established between these towns with the installation of a large cable carrier telephone system for the G.P.O. This has been supplied and commissioned by A.T.E.

This system, operating on twin low capacitance cables, carries 60 channels in the 12-252 Kc/s frequency band on each of the 7 equipped pairs.

The repeaters are fed with power along the cable, eliminating the need for expensive power plant at intermediate stations.

THIS IS THE FIRST INSTALLATION OF ITS KIND IN THE UNITED KINGDOM



Suite of A.T.E. baysides at the Middlesbrough terminal station

The link between Leeds and Middlesbrough has repeaters in the line at Thorpefield Boroughbridge, Jeater Houses, Walshford Bridge, Scarcroft and Stokesley.

KS

s apart





AUTOMATIC TELEPHONE & ELECTRIC COMPANY LTD.

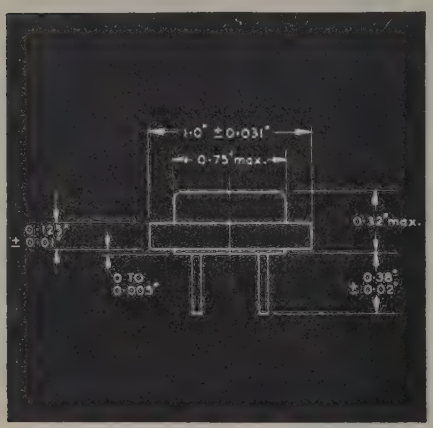
STROWGER HOUSE, ARUNDEL STREET
LONDON, W.C.2.
Telephone: TEMple Bar 9262
Telegrams: Strowger, Estrand, London

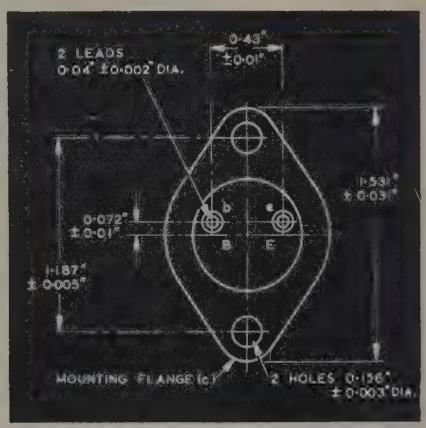
AT 5091

Audio power output transistors

TYPES XC141 and XC142

These germanium p-n-p alloy junction transistors are designed for use in Class A and Class B power output stages of audio frequency amplifiers. Full particulars of these and other Ediswan Mazda semiconductor devices will be sent gladly on request. If you wish to be kept up to date with the latest developments in this field, please ask us to add your name to our semiconductor mailing list.





MAXIMUM RATINGS (Absolute Values)	XC141	XC142
Peak collector to base voltage (volts)	−40	−60
Peak collector to emitter voltage, emitter non-conducting (volts)	−40	−60
Peak collector to emitter voltage, emitter conducting (volts)	−32	−32
D.C. Emitter to base voltage (volts)	−12	−12
Peak collector current (amps)	− 3·0	−3·0
D.C. Collector current (amps)	− 1·5	−1·5
Collector dissipation (mounting flange temperature 80°C) (watts)	11	11

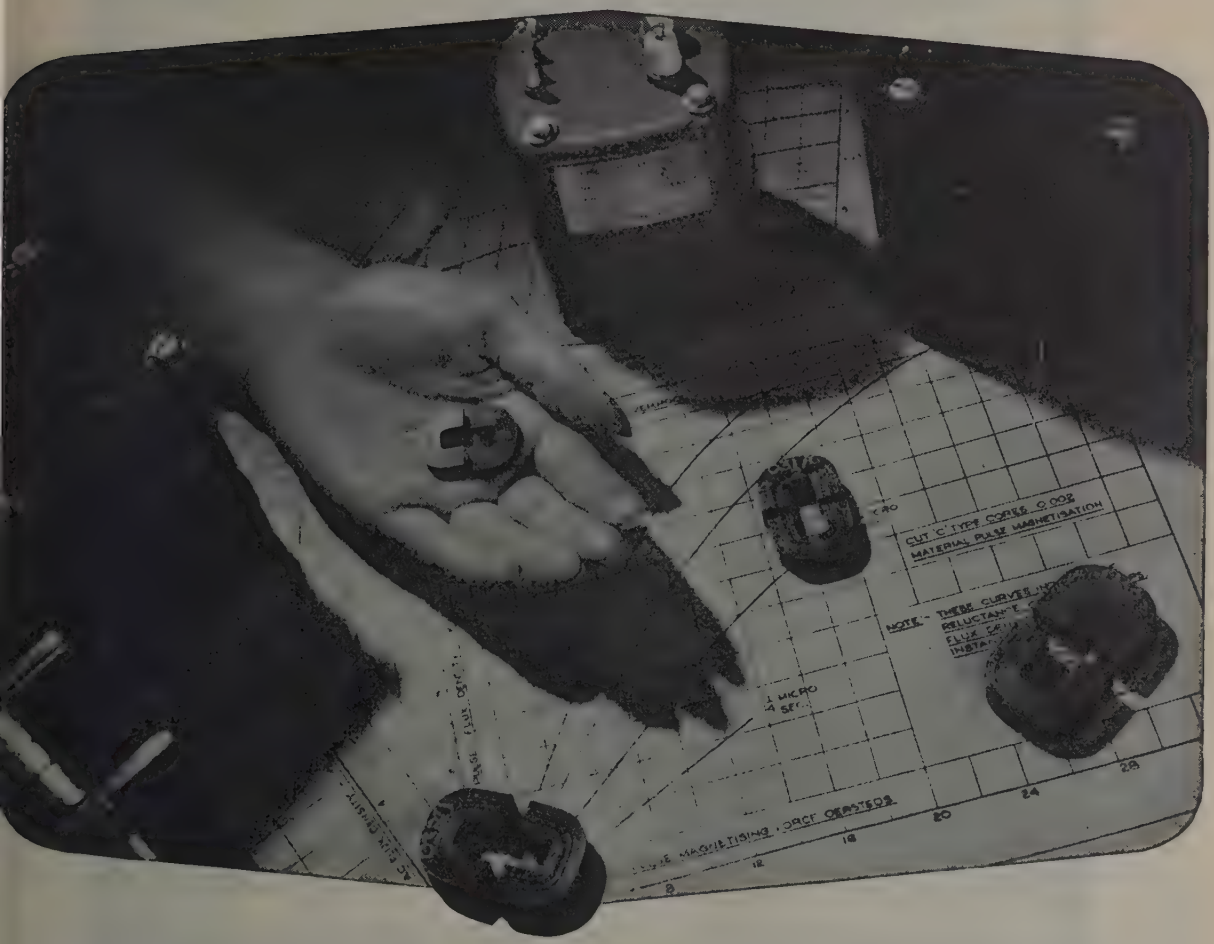
EDISWAN SEMICONDUCTORS
MAZDA

Associated Electrical Industries Ltd
Radio and Electronic Components Division
PD 15, 155 Charing Cross Road, London, W.C.2
Tel: GERrard 8660 Telegrams: Sieswan Westcent London

CRC 15/59

'ENGLISH ELECTRIC'

'C' CORES



for smaller, lighter transformers and chokes 'ENGLISH ELECTRIC' 'C' type transformer cores are made from high-quality cold-rolled grain oriented silicon steel strip. A wide range of cores is available and these have been developed to meet the exacting standards of the electronic industry, for smaller, lighter and more efficient transformers and chokes.

Some advantages of using ENGLISH ELECTRIC *'C' cores:—*

Easy to store and handle, their use results in reduced assembly time.
Transformers are up to 30% smaller and lighter.
Transformers have lower losses, lower magnetising VA and lower leakage flux.
Cores are suitable for oil-filled, open-type or resin-cast transformers.
Cores are stable up to 250°C.
Cores are available from stock.

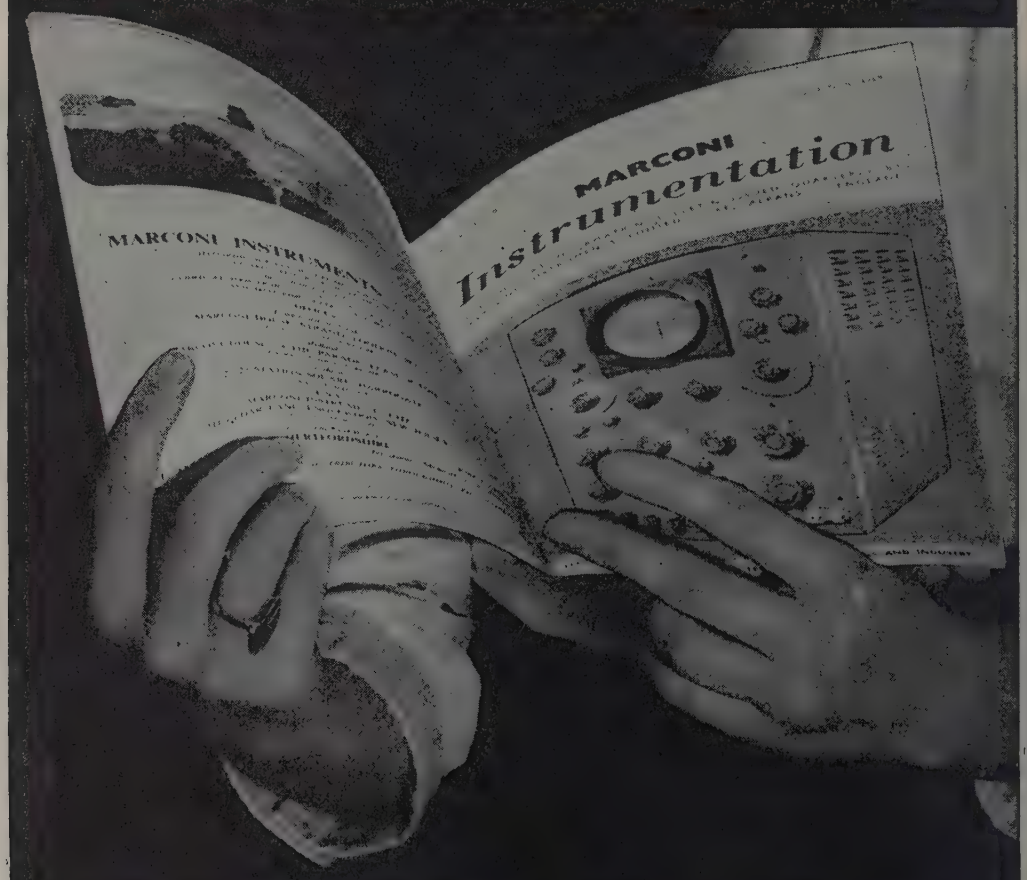
full details from

The English Electric Company Limited, Transformer Sales & Contracts Dept.,
East Lancashire Road, Liverpool, 10. Telephone: AINtree 3641

THE ENGLISH ELECTRIC COMPANY LIMITED, MARCONI HOUSE, STRAND, LONDON, W.C.2.
WORKS: STAFFORD · PRESTON · RUGBY · BRADFORD · LIVERPOOL · ACCRINGTON

FL.39

MARCONI
Instrumentation



A FREE QUARTERLY BULLETIN OF TECHNICAL PROGRESS

MARCONI INSTRUMENTS

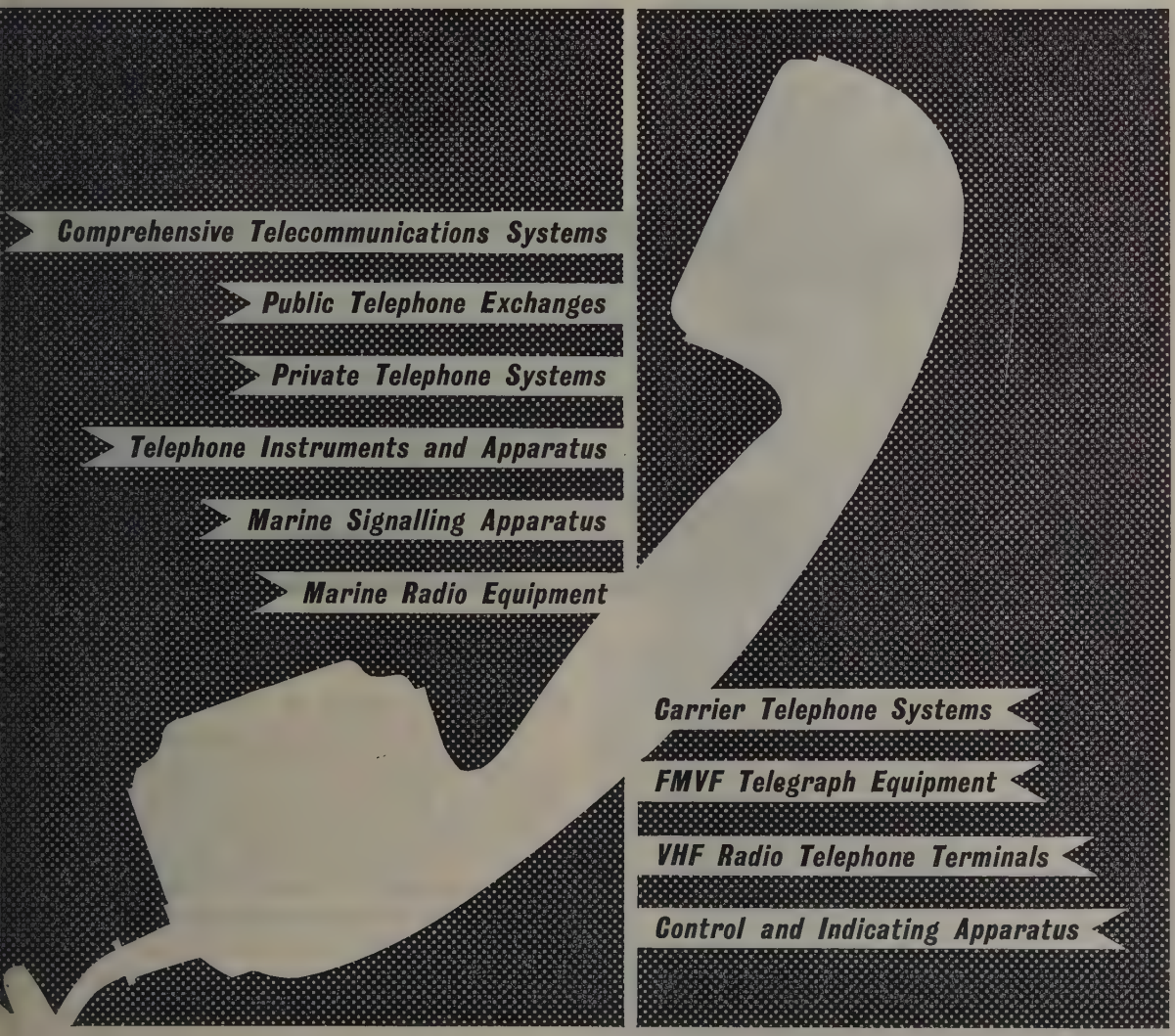
Marconi Instrumentation deals primarily with trends and developments in telecommunication test equipment, and provides current data relating to instrumentation techniques. Written by specialists, it is an authoritative and up-to-date record of progress in this important field, and if you are not already receiving a free copy every quarter, please let us have your name and address.

Please address enquiries to MARCONI INSTRUMENTS LTD., at your nearest office:
London and the South: Marconi House, Strand, London, W.C.2. Telephone: COVent Garden 1234
Midlands: Marconi House, 24 The Parade, Leamington Spa. Telephone: 1408. *North:* 23/25 Station Square, Harrogate. Telephone: 67455
Export Department: Marconi Instruments Ltd., St. Albans, Herts. Telephone: St. Albans 56161

TC158

AEI

Telecommunications Division



Associated Electrical Industries Limited

Telecommunications Division

WOOLWICH, LONDON, S.E.18

SIEMENS EDISWAN

formerly the Telecommunications Division of

SIEMENS EDISON SWAN LIMITED, WOOLWICH, LONDON, S.E.18

SIEMENS EDISWAN

TA3382



Marconi Crystals for stability and precision

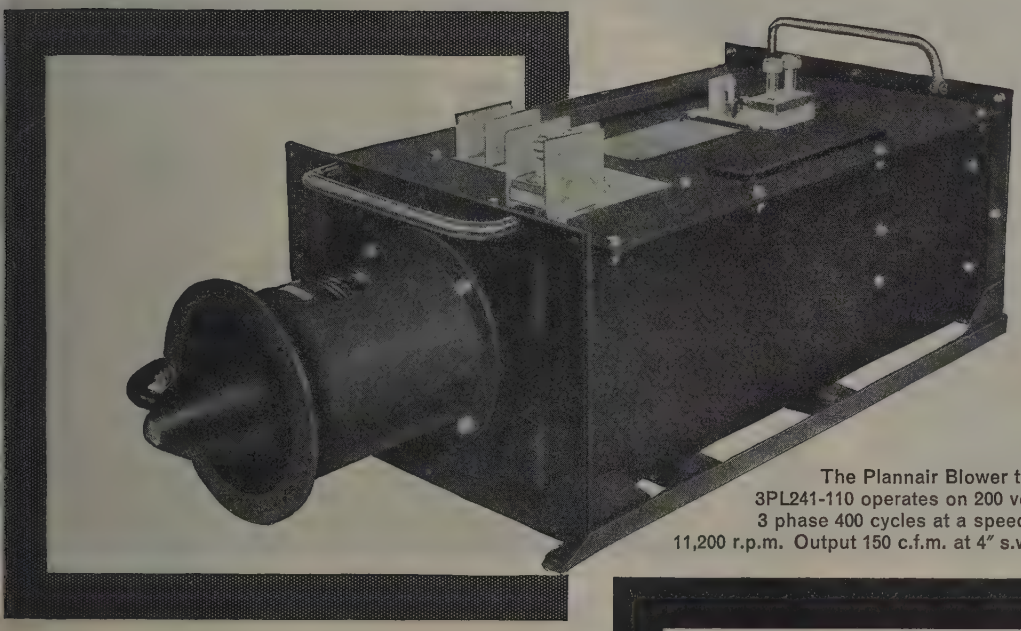
The experience gained in manufacturing quartz crystals to the stringent requirements of our own apparatus and of Service equipment enables us to offer a comprehensive range of crystals covering the frequency band 1.0 Kc/s to 62 mc/s at extremely competitive prices.

Years of intensive research and development work guarantee the reliability and quality of this Marconi product.

MARCONI **CRYSTALS FOR ELECTRONICS**

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND

PLANNAIR keeps things temperate for Ferranti aboard the **'Vanguard'**



The Plannair Blower type 3PL241-110 operates on 200 volts 3 phase 400 cycles at a speed of 11,200 r.p.m. Output 150 c.f.m. at 4″ s.w.g.

This beautifully engineered transformer/rectifier unit by Ferranti is installed in Britain's newest airliner, the Vickers Vanguard. The unit is, in Ferranti's words, ' an assembly giving the minimum overall dimension and weight.'

In such compact equipments, cooling is a critical factor. It must be adequate and the cooling source must also be lightweight and efficient. Like Ferranti, most manufacturers of advanced electronics rely on Plannair blowers to solve temperature control problems.

On a weight-to-output basis, Plannair blowers—the whole family of them—are the most efficient in the world.

Whenever you are faced with the problems of getting enough air in the right place, take them to Plannair—that's what the rest of the industry does!



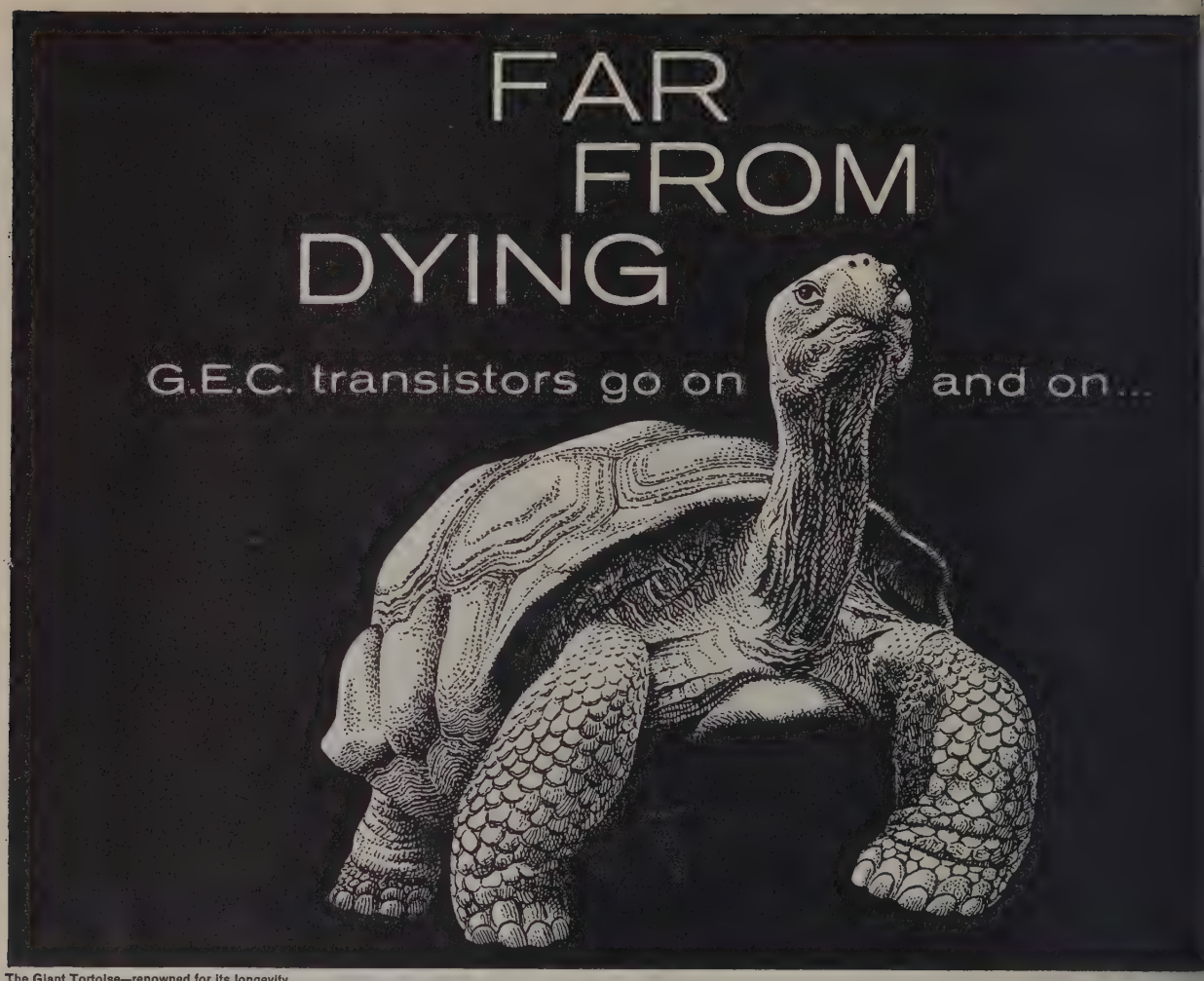
Illustration shows four single stage blower units type 3PL241-110 installed for transformer/rectifier cooling on the main d.c. power supply unit in the Vickers Vanguard.

PLANNAIR LIMITED
Windfield House • Leatherhead • Surrey
Tel: Leatherhead 4091/3 & 2231

PLA32

FAR FROM DYING

G.E.C. transistors go on and on...



The Giant Tortoise—renowned for its longevity.

GET103 Failure Rate is only 0.06% in 1,000 hours!

5,000 GET 103 transistors—random samples from regular production—have been submitted to electrical life test. Up to the end of 1959 only three catastrophic (inoperative) failures had occurred, indicating a failure rate of 0.06% per 1,000 hours. Some of these tests continue to run indefinitely, and we are thus building up life information for periods of many thousands of hours.

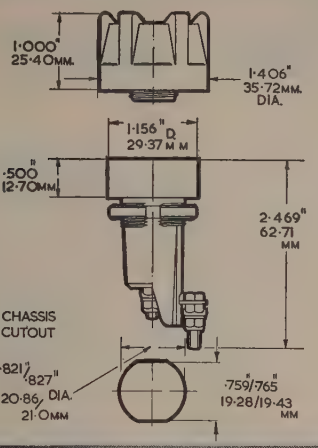
Transistors can also show changes in electrical characteristics during life: gain and leakage currents are the parameters most likely to change, but the precise operating conditions determine the extent of the changes. Our life tests show that after an initial "settling-down" period, the subsequent rate of change in the characteristics of G.E.C. transistors is extremely small, indicating that they will continue functioning satisfactorily for tens of thousands of hours.

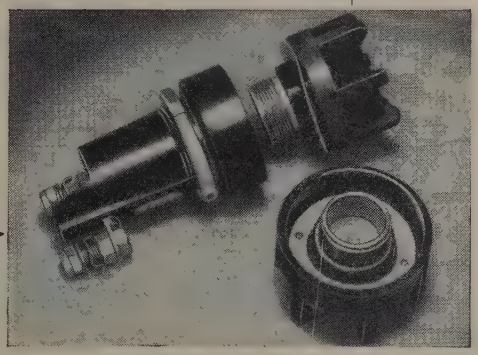
G.E.C.

SEMICONDUCTORS

For full information on these and many other types, please write to:—
G.E.C. Semiconductor Division, School Street, Hazel Grove, Stockport, Cheshire.
Tel: Stepping Hill 3811 or for London Area ring Temple Bar 8000, Ext. 10.

"BELLING-LEE" Fuseholders





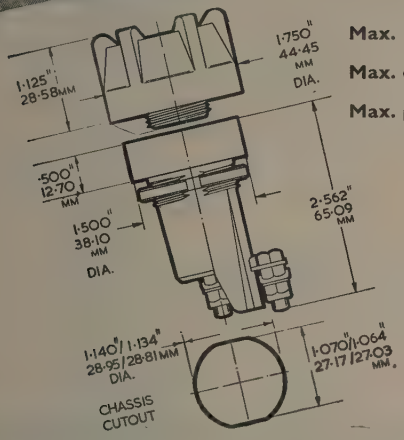


L.1302 PANEL FUSEHOLDER. Size 1 (with large lid)
L.1302/L. Lid only (large)
L.1302/B. Body (common)

L.1303. PANEL FUSEHOLDER. Size 1 (with small lid)
L.1303/L. Lid only (small)
L.1302/B. Body (common)

L.1304. PANEL FUSEHOLDER. Size 2 (with large lid)
L.1304/L. Lid only (large)
L.1304/B. Body (common)

L.1305. PANEL FUSEHOLDER. Size 2 (with small lid)
L.1305/L. Lid only (small)
L.1304/B. Body (common)



Max. working volts: L.1302: 440 V. L.1303: 250 V. peak.
Max. current rating: Both types: 15 amp.
Max. panel thickness: Both types: 7/32 in.

Max. working volts: L.1304: 440 V. L.1305: 250 V. peak.
Max. current rating: Both types: 30 amp.
Max. panel thickness: 7/32 in.

This series of heavy duty panel mounting H.R.C. fuseholders was developed in co-operation with the Ministry of Supply for use in military aircraft and vehicles.

In these clamp type fuseholders, fuse withdrawal is effected automatically as the lid is unscrewed.

Each of the two large versions provides fingerguard protection as an additional safety feature.

Mounting to the panel is made through a single hole shaped to engage the anti-swivel flats in the holder, which is secured with a ringnut. All terminals are complete with washers and shakeproof nuts.

Most "Belling-Lee" products are covered by patents or registered designs or applications.

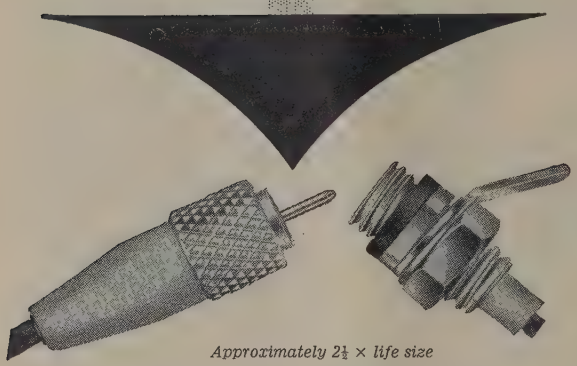
Regd

TERMINALS . PLUGS &
SOCKETS . CIRCUIT
PROTECTION DEVICES
INTERFERENCE FILTERS
RECEIVING AERIALS

BELLING & LEE LTD
GREAT CAMBRIDGE ROAD, ENFIELD, MIDDX., ENGLAND
Telephone: Enfield 3322 · Telegrams: Radiobel, Enfield

The very best connections ...secured by Plessey

SUB-MINIATURE COAXIAL CONNECTORS



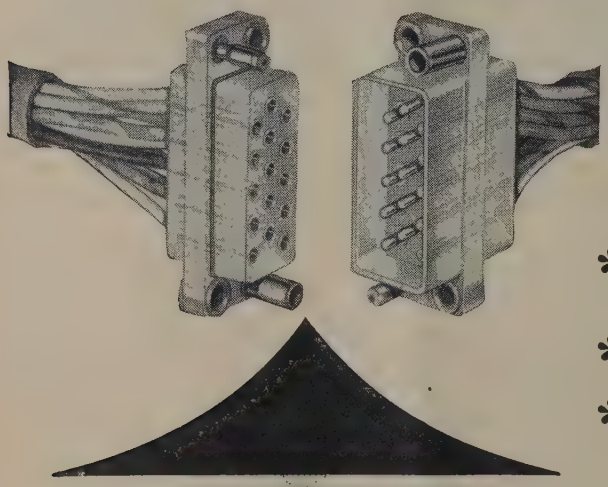
Approximately 2½ × life size

As a contribution towards increasingly compact equipment, Plessey have introduced this new, highest quality and fully comprehensive range to allow a new approach on applications hitherto restricted by the limitations of existing connectors.

Designed for the matched impedance coupling of high frequency coaxial cables operating in the super high frequency bands, these connectors—

* have a working voltage of 650 volts PEAK at sea level, and matched impedance coupling of 50 ohm lines is accommodated.
* have hard gold plated contacts on silver plate to give maximum performance with minimum voltage drop.

SERIES '110' (15- and 30-way) MINIATURE RECTANGULAR CONNECTORS



Developed specifically by Plessey to meet the demand for a safe, inexpensive connector for commercial applications, this new series embodies excellent electrical and mechanical characteristics, and the many unique features that make it really outstanding include:—

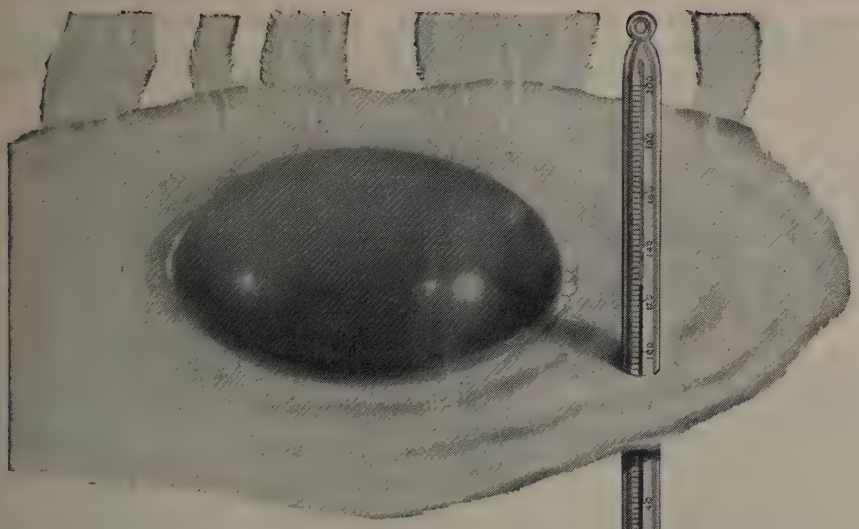
* Plug pins and socket inserts are polythene shrouded to dispense with gaskets and ensure insert anchorage.
* Mismating is prevented by corner pins and corner sockets.
* Extreme simplicity of wiring, demands less-skilled operation than the orthodox methods of soldering pins *in situ*.

For further information please write for Publication Nos. 128 and 114

Plessey

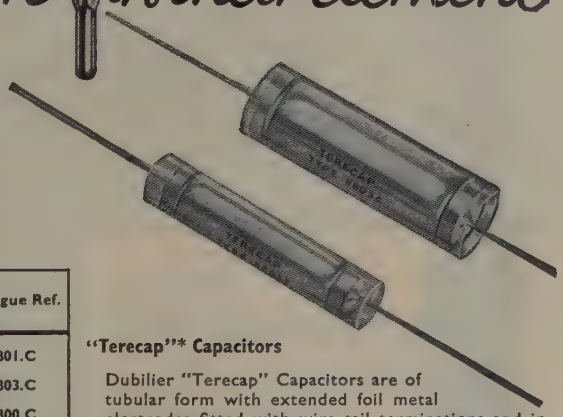
WIRING & CONNECTORS DIVISION · THE PLESSEY COMPANY LIMITED
CHENEY MANOR · SWINDON · WILTS · TELEPHONE: SWINDON 6251

Overseas Sales Organisation: PLESSEY INTERNATIONAL LIMITED · ILFORD · ESSEX



When it's hot enough to fry an egg

"Terecaps" are in their element



SIZES AND RATINGS

Capacitance	Working Voltage d.c. at 71°C	Working Voltage d.c. at 125°C	Test Volts d.c.	Dimensions ins. Length	Dimensions ins. Dia.	Catalogue Ref.
0.1	150	125	300	1¼	½	8801.C
0.25	150	125	300	1⅞	½	S-8803.C
0.5	150	125	300	1⅞	⅝	S-8800.C
1.0	150	125	300	1⅞	¾	S-8804.C
0.1	250	180	500	1¼	½	8801.C
0.25	250	180	500	2⅛	½	8803.C
1.0	250	180	500	2⅛	¾	8804.C
0.1	350	250	700	1⅜	½	8802.C
0.25	350	250	700	1⅞	¾	S-8804.C
1.0	350	250	700	2½	1	8806.C

Other capacitance values can be supplied to order. We invite your enquiries.

"Terecap"* Capacitors

Dubilier "Terecap" Capacitors are of tubular form with extended foil metal electrodes fitted with wire tail terminations and incorporate a non-hygroscopic film dielectric. Being designed to meet abnormal atmospheric conditions such as obtain in tropical zones the capacitors are supplied hermetically sealed in metal containers with ceramic end-seals.

* *A Registered Dubilier Trade Mark.*

Dubilier "Terecap" Capacitors have these outstanding features:—

1. Can be used up to 125°C. with voltage de-rating above 70°C.
2. High insulation resistance, more than twenty times that of paper dielectric capacitors. (10,000ΩF at 20°C.)
3. Compactness.
4. Excellent capacitance stability over a wide temperature range. (Normal capacitance tolerance ± 20%)
5. Power Factor 0.5% at 20°C. for 1 kc/s.

DUBILIER

DUBILIER CONDENSER CO. (1925) LTD., DUCON WORKS, NORTH ACTON, LONDON, W.3.
Phone : ACOrn 2241

DN.225A

OUR **TRAVELLING WAVE TUBES** UNLIKE THE ANGEL FISH, ARE NOT AS BROAD AS THEY'RE LONG BUT THEY DO HAVE EXTREMELY WIDE BAND WIDTH!





We have a wide selection of Low Noise Tubes, Intermediate (medium power) Tubes and Power Tubes, covering many frequency bands. For further details and information concerning any of our range of thermionic tubes for industry, write to the Company.

'ENGLISH ELECTRIC'

AGENTS THROUGHOUT THE WORLD

STAND E215

EXHIBITION

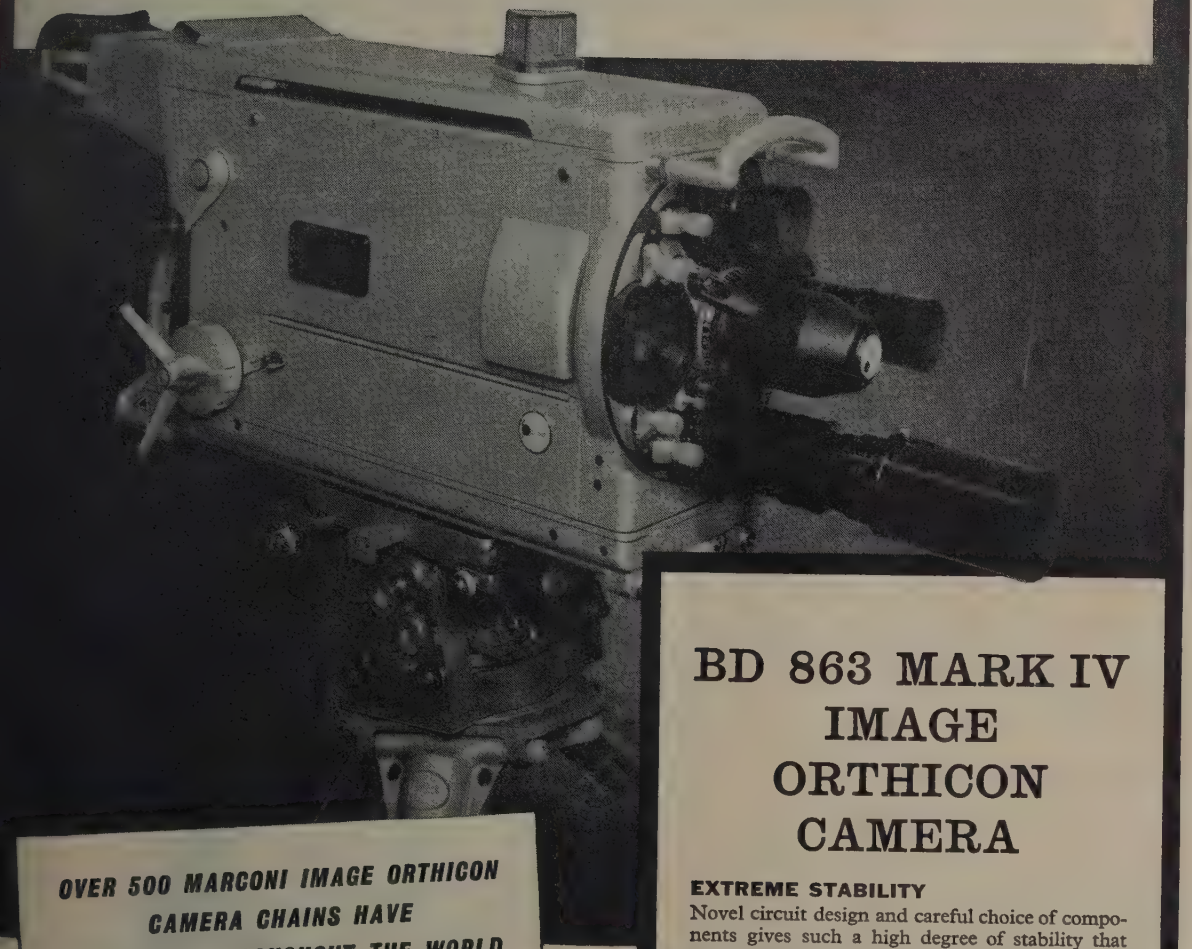
ENGLISH ELECTRIC VALVE CO. LTD.

Chelmsford, England
Telephone: Chelmsford 3491

THE MARK IV CAMERA CHAIN

EXPERIENCE COUNTS

Marconi's pioneered the use of the 4½ inch Image Orthicon Camera using the tube developed by their associates, the English Electric Valve Company. Marconi's have amassed more 'know-how' on the use of the 4½ inch Image Orthicon than any other manufacturer.



OVER 500 MARCONI IMAGE ORTHICON CAMERA CHAINS HAVE BEEN SOLD THROUGHOUT THE WORLD

BD 863 MARK IV IMAGE ORTHICON CAMERA

EXTREME STABILITY
Novel circuit design and careful choice of components gives such a high degree of stability that operational controls have been removed from the camera.

FIRST CLASS PICTURE QUALITY
The 4½ inch Image Orthicon tube gives a picture quality substantially better than any other type or size.

LIGHT AND COMPACT
By reducing and simplifying the camera electronics its weight has been held below 100 lb. and its size made correspondingly small.

MARCONI

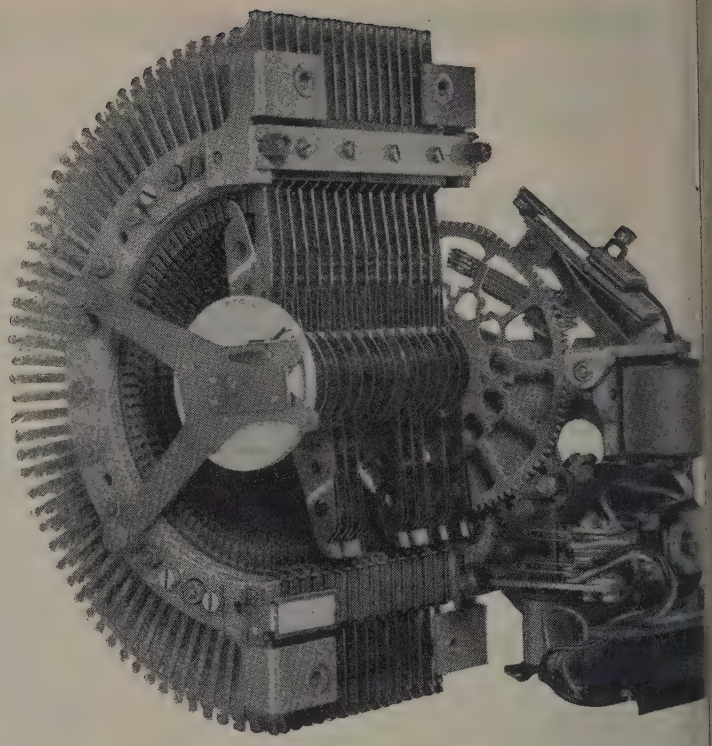
COMPLETE SOUND AND TELEVISION SYSTEMS

RCONI'S WIRELESS TELEGRAPH COMPANY
MITED · CHELMSFORD · ESSEX · ENGLAND

M4D

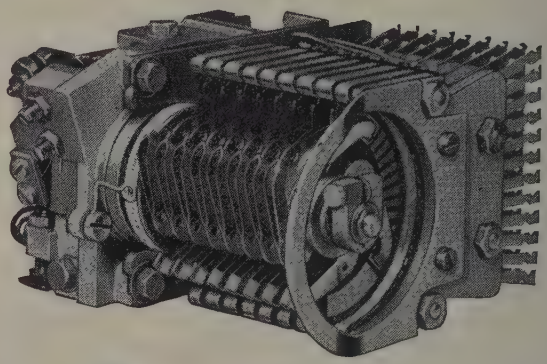
These 3 outstanding uniselectors

provide long life and reliability in operation



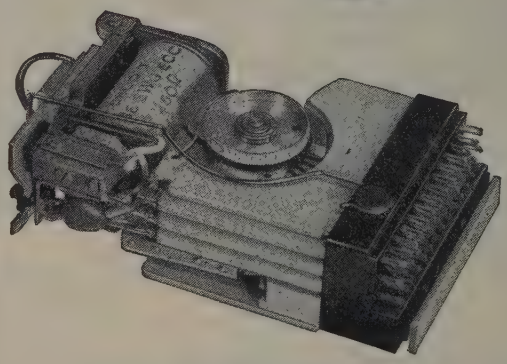
The high speed motor uniselector

has an availability of 52 outlets in a half revolution with up to 16 parallel channels, and is capable of high speed search if required. It is also suitable for scanning and other collection functions, or for the routing of information to a multitude of points. It can be brought to rest at any desired outlet by means of simple circuitry, and a homing drive is provided so that it can be readily restored to the normal position. Arrangements can also be made for step-by-step working.



The digit switch

is specifically designed for step-by-step operation in response to impulses, and has an availability of 12 outlets with 8 parallel channels. It may be used to set up a marking condition for the control of the Motor Uniselector.



The miniature uniselector

is a compact switch of unique design which occupies no more space than a standard 3000-type relay. It has an availability of 12 outlets with 3 parallel channels. A great advantage is that it incorporates a plug-in type of mounting so that the entire device can be removed and replaced without detaching any soldered connections.

AEI

Full technical information will gladly be supplied on application

ASSOCIATED ELECTRICAL INDUSTRIES LTD.
Telecommunications Division,
Telephone Apparatus Department,
Woolwich, London S.E.18.
Woolwich 2020 Ext. 326

Printed Circuit Counter Panels



A complete range of transistorized counter panels of common size, fixing method and electrical connexion, designed to provide a flexible unit system whereby any special requirements in the counting or data processing fields can be quickly built up.

A fully illustrated brochure giving complete performance and specification figures for every panel in the range is available on request.

50kc/s Scaler
1Mc/s Scaler
Input Amplifier
Gate Unit
10kc/s Oscillator
1Mc/s Oscillator
Power Unit
50kc/s Read-out Scaler
1Mc/s Read-out Scaler
4 Channel Output Unit
Read-out Unit
Meter Display Unit
Lamp Display Unit
Numerical Indicator Tube
Shift Register Stage
Shift Register Driver

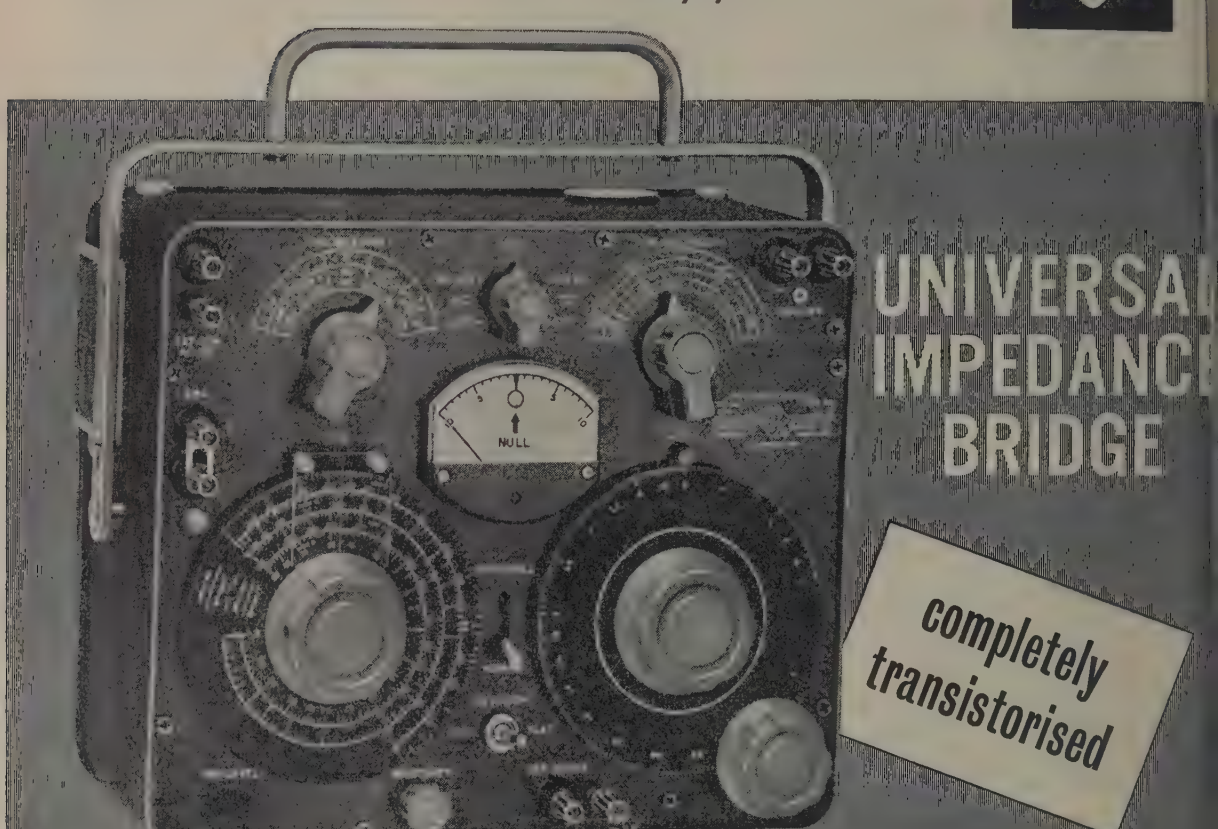
CINTEL

RANK CINTEL LIMITED
Worsley Bridge Road · London · SE 26
HITher Green 4600

Sales and Servicing Agents: Atkins, Robertson & Whiteford Ltd. Industrial Estate, Thornliebank, Glasgow; McKellen Automation Ltd., 122 Seymour Grove, Old Trafford, Manchester, 16; Hawnt & Co. Ltd., 59 Moor St. Birmingham, 4.

The Most Significant Advance in Bridge Design

in many years



'GENERAL RADIO'
Type 1650 -A
Universal Impedance Bridge

With 'Orthornull' to speed balance — eliminates sliding null.

Wide Ranges — **R** 0·001Ω to 10MΩ **L** 1μh to 1000h
C 1μμf to 1000μf **D** 0·001 to 50 at 1 kc/s
Q 0·02 to 1000 at 1 kc/s

Basic 1% Accuracy for R, L, C — 5% for D and Q. Not reduced at range extremes

Ready to use — battery operated with built-in null detector and 1-kc/s generator (20c/s to 20 kc/s frequency range with external generator)

Unique Carrying Case allows panel to be tilted to any convenient angle . . . closes for complete protection

Write for complete information

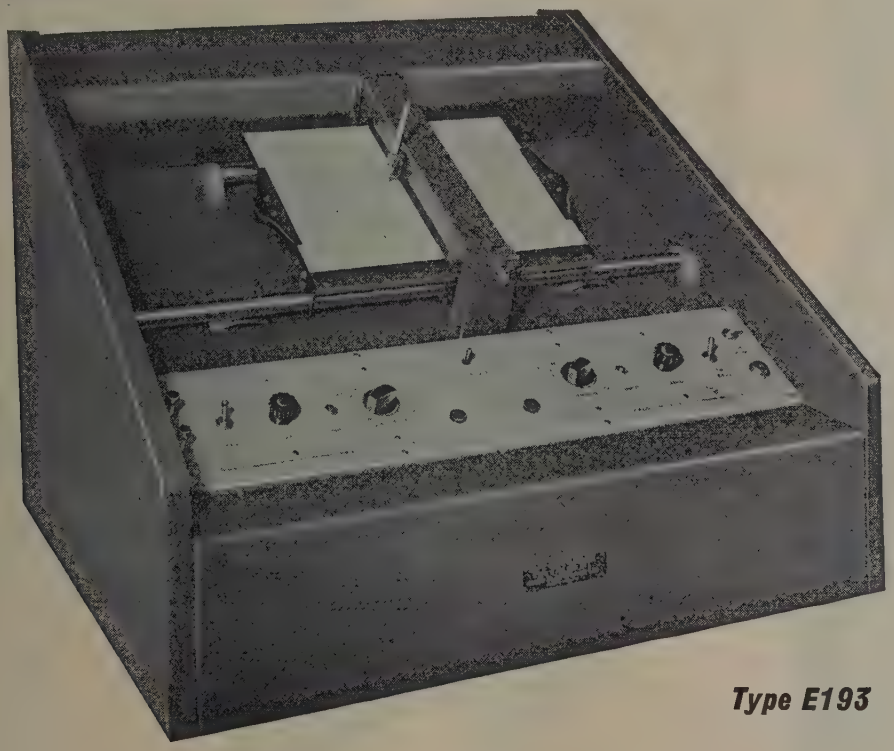
Claude Lyons Ltd.

Sole U.K. representatives of the General Radio Company, U.S.A.

Valley Works · Hoddesdon · Herts · Telephone: Hoddesdon 4541 (6 lines)

CL 53

EKCO X-Y Recorder



Type E193

Ekco Type E193 is a high-speed X-Y Recorder which plots the relationship between two inter-dependent variables directly on to standard 10″ x 7½″ graph paper.

The wide sensitivity range (1 mV. per 1″ to 50 V. per 1″) and exceptionally fast response (full-scale travel in 200 milli-seconds) make it unusually versatile in the fields of performance investigation and process control.

This instrument will find wide use in research laboratories, industrial organisations and hospitals or in any field in which it is necessary to chart rapidly-changing parameters accurately.

Sales, Installation and Service

EKCO ELECTRONICS LIMITED

SOUTHEND-ON-SEA · ESSEX · Tel. Southend 49491

WPS 306

VHF

G.E.C.

Radio Telecommunicati



5 circuit V.H.F
junction radio
terminal equipment.

Over 1000 5-circuit V.H.F. Junction Radio Equipments have been supplied to 28 countries throughout the world.

The vast experience gained by G.E.C. in surveying, planning, designing, ma

G.E.C.

EVERYTHIN

THE GENERAL ELECTRIC COMPANY LIMITED

WORKS AT COVENTRY,

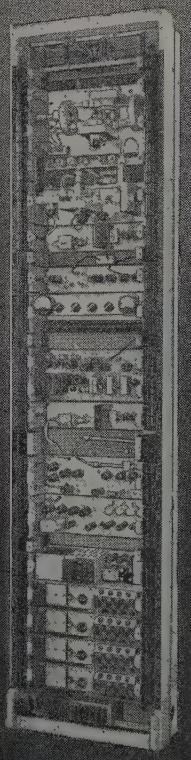


UHF

uipment

throughout the world

Microwave systems providing over 7000 wide band channel miles for television or multi-circuit telephony have been installed or ordered for 7 countries throughout the world.



240-circuit U.H.F. radio terminal equipment (less covers)

ling radio systems in all five continents is available to prospective customers.

MUNICATIONS

TELEPHONE WORKS, COVENTRY, ENGLAND

ROUGH, PORTSMOUTH

G.E.C.

G.E.C 24

PHILIPS

In Science and Industry alike . . .

among technicians, manufacturers and those engaged in the sale of electrical products — as well as among the public at large, the Philips emblem is accepted throughout the World as a symbol of quality and dependability.

PHILIPS ELECTRICAL LTD

Century House · Shaftesbury Avenue · London · WC2

Radio & Television Receivers · Radiograms & Record Players · Stereo & Hi-Fi Equipment · Tape Recorders · Dictation Machines · Gramophone Records · Tungsten, Fluorescent, Blended & Discharge Lamps & Lighting Equipment · " Photoflux " Flash Bulbs · " Philishave " Electric Shavers · Health Lamps · Hearing Aids · Electric Blankets · Domestic Electrical Appliances · X-Ray Equipment · High Frequency Heating Generators · Electro-Medical Apparatus · Arc Welding Plant & Electrodes · Electronic Measuring Instruments · Magnetic Filters · Sound Amplifying Installations · Cinema Projectors

What every Engineer should know.... about

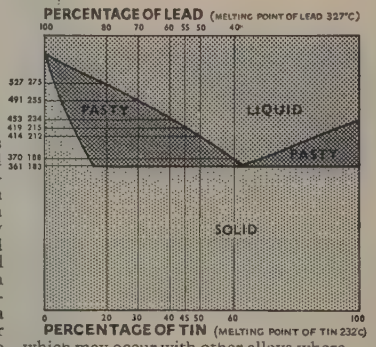
MULTICORE solder



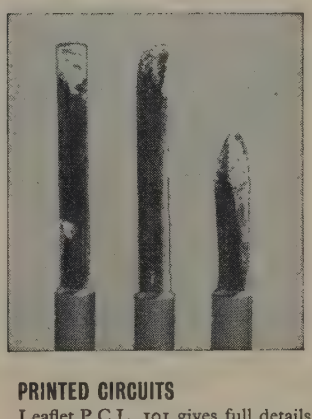
ERSIN MULTICORE 5-CORE SOLDER

The A.I.D. approved type 362 flux, incorporated in Ersin Multicore 5-Core Solder, is very effective on heavily oxidised surfaces and often allows the use of a lower tin content alloy. Ersin Multicore 5-Core Solder is supplied on 7 lb. reels in 9 standard gauges and 6 alloys. Even gauges from 24-34 s.w.g. are available in 2 alloys on 1 lb. and ½ lb. reels.

CONSTITUTION OF ALLOYS OF ERSIN MULTICORE SOLDER



The diagram shows that all the standard alloys of Ersin Multicore Solder have a plastic range, i.e., on heating they are pasty between the solid and liquid states. Practical experience has shown that there are advantages in having a plastic range, e.g. for tag jointing where the use of 60/40 alloy obviates fractures which may occur with other alloys where there is slight vibration while the solder is setting solid.



PRINTED CIRCUITS

Leaflet P.C.L. 101 gives full details of a complete soldering process developed by Multicore Laboratories for printed circuits.

SAVBIT TYPE 1 ALLOY *Made under sole British Licence of Patent No. 721,881.*

Savbit Type 1 alloy was developed after extensive research in the Multicore Laboratories into the main causes of bit wear. It incorporates a small percentage of copper which prevents absorption of copper from the bit into the solder alloy. After prolonged tests, it was found that the life of solder bits was increased by up to 10 times. The speed of soldering is not affected.

SPECIAL ALLOYS

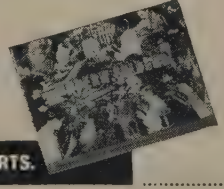
4 special alloys can be supplied for particular purposes:

Comsol with a high melting point of 296°C.

T.L.C. alloy with a melting point of 145°C is made from tin, lead and cadmium.

P.T. (Pure Tin) alloy, with a melting point of 232°C, is lead-free.

L.M.P. alloy, with 2% silver content, which melts at 179°C for silver-coated components.



PUBLICATIONS

Laboratory engineers and technicians are invited to write on their company's letter-heading for the latest edition of Modern Solders. It contains data on melting points, gauges, alloys, etc.

MULTICORE SOLDERS LTD., MULTICORE WORKS, HEMEL HEMPSTEAD, HERTS.

Tel: Boxmoor 3636 (4 lines). Grams and Cables: Multicore Hemelstead.

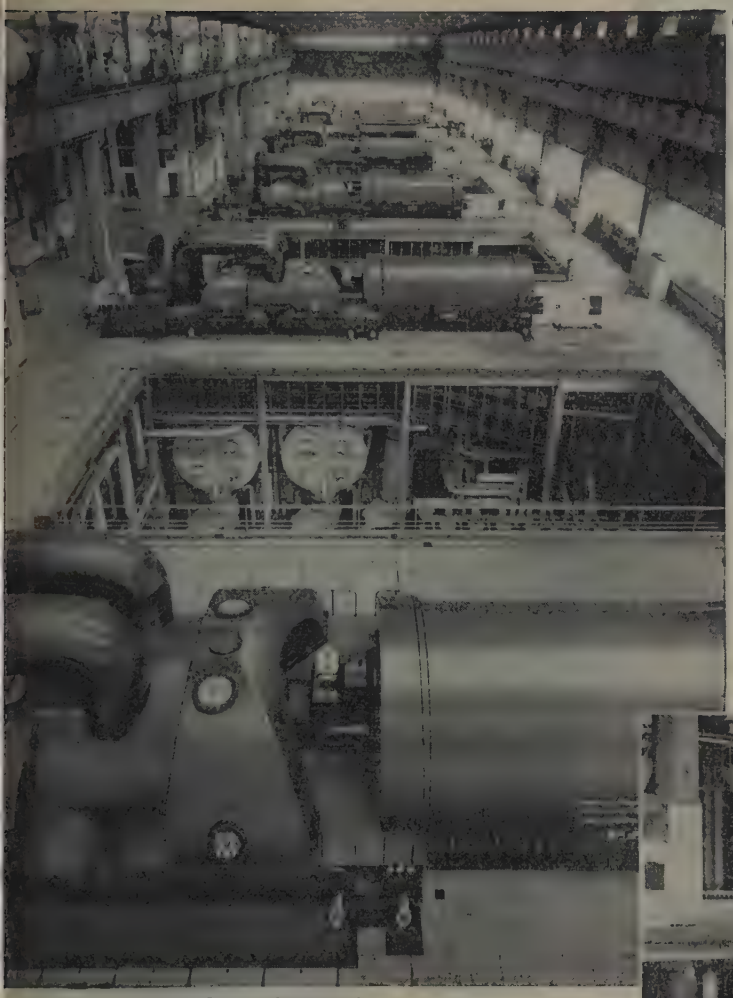
CASTLE DONINGTON

C.E.G.B. East Midlands Division.

The largest single thermal power station in Western Europe

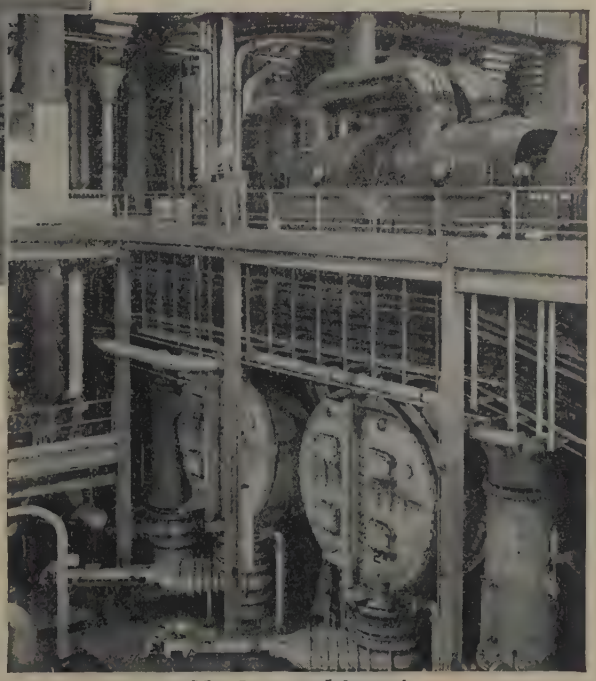
Castle Donington houses six 100 MW 3000 rpm steam turbine-generators together with their associated condensing plants and feedwater heating equipments. All this power plant was manufactured by A.E.I. Turbine-Generator Division.

Castle Donington recorded the highest average overall thermal efficiency of any British power station during the three years ending 31st March 1959.



General view of the complete installation of six 100 MW turbine-generators

The two-cylinder, close-coupled turbines operate with steam conditions at the turbine stop valve of 1500 psig, 1050°F and exhaust to A.E.I. condensers.
Direct-coupled generators produce 3 phase, 50 c/s alternating current at 13·8 kV. The rotor windings are directly cooled by hydrogen circulating in contact with the copper.
The generators are directly connected to the A.E.I. main step up transformers and the power is controlled to the grid by A.E.I. circuit-breakers.



No. 2 set and its twin-shell A.E.I. Condenser.

AEI

Associated Electrical Industries Limited

TURBINE-GENERATOR DIVISION

Enquiries to: AEI Turbine-Generator Division; Trafford Park, Manchester 17, or your local AEI office.

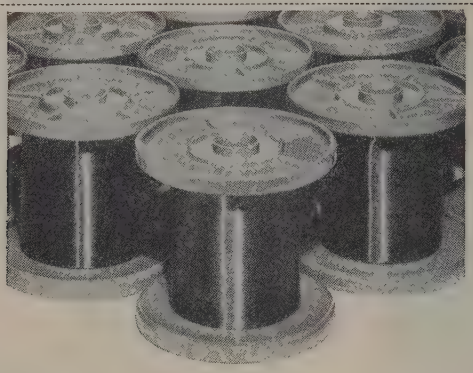
WORKS AT MANCHESTER AND RUGBY, ENGLAND · GLASGOW, SCOTLAND · LARNE, NORTHERN IRELAND

B/A 907

LEWCOS

Supplied with standard coverings of cotton, silk, rayon, enamel, LEWMEX (synthetic enamel) glass and abestos.

Insulated resistance wires



for resistors in instruments, radio, television, electronic and control apparatus and many other applications.

★ *Available over a range of sizes in VACROM (nickel-chrome) 80/20 or 15% and EUREKA (cupro-nickel). All conform to BSS 115/1954.*

A technical information and advisory service is offered.

THE LONDON ELECTRIC WIRE COMPANY AND SMITHS LIMITED

LEYTON · LONDON · E10

ADCOLA (Regd. Trade Mark)

Soldering Instruments

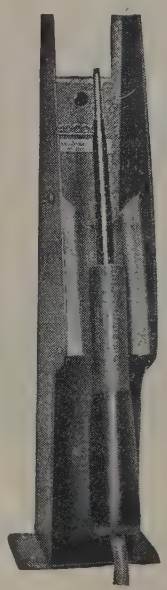
ILLUSTRATED

PROTECTIVE SHIELD (CAT. No. 68)

$\frac{1}{8}$ in. BIT MODEL (CAT. No. 70)

Primarily developed for the TRANSISTOR & ELECTRONIC ERA.

Possessing the sharp heat essential for the quick jointing of Transistors, Resistors, etc., thereby avoiding damage to the equipment from heat transference



Cover all requirements for thorough solder jointing in all the fields of

TELECOMMUNICATIONS

Fully Insulated Elements

Suited to daily use for bench line production

MANUFACTURED IN ALL VOLT RANGES

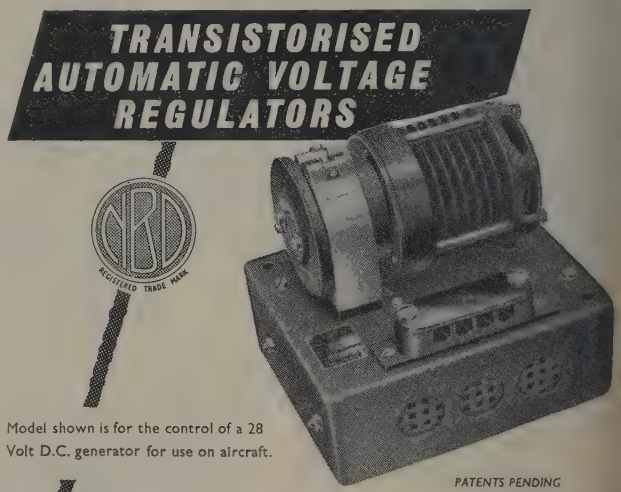
British and Foreign Pats. Reg. designs, etc.

For further information apply Head Office:

ADCOLA PRODUCTS LTD.
GAUDEN ROAD
CLAPHAM
HIGH STREET
LONDON, S.W.4

Tel.: MACaulay 4272 & 3101

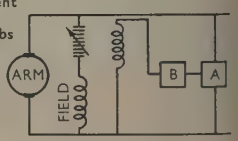
TRANSISTORISED AUTOMATIC VOLTAGE REGULATORS



Model shown is for the control of a 28 Volt D.C. generator for use on aircraft.

PATENTS PENDING

Regulation closer than ± 1% between extremes of temperature from −60°C to +70°C
Speed of response 50/60 milliseconds.
For industrial purposes at normal ambient temperatures regulation within ± 0·5%.
Dimensions 5" × 6" × 5¼" high. Weight 4lbs



A=REFERENCE BRIDGE
B=TRANSISTOR AMPLIFIER

NEWTON DERBY

NEWTON BROS. (DERBY) LTD.
ALFRETON ROAD · DERBY
PHONE: DERBY 47676 (4 LINES) GRAMS: DYNAMO, DERBY
London Office: IMPERIAL BUILDINGS, 56 KINGSWAY, W.C.2

we can get you out of a wide band problem



And if you're in a rage over rectifiers or in a diode of despair—don't worry. We'll get you out of that too. (Why, we've made some of our customers actually beam at tetrodes!) It's because we've experienced over 40 years bottling up valves—Tx, Rx, rf, audio, hard, soft, gov't., special—that problems like this no longer hold any terrors for us. Also, we've everything on the research and production side you could possibly wish for. So the next time you're enveloped in a valve problem, let us know.

TYPE	g_m (mA/V)	C_a—all less gl (pF)	C_{gl}—all less a (pF)	P_a (W)
Z319	19	3	8	2.5
Z759	15	3	13	5
417A	25	0.35	6.5	4
R5559	25	5	12	4.5
A2674	32	3.5	15	4.5

Many other types, for all applications—Let us know what you're interested in and we will be pleased to send you details.

G.E.C. valves are obtainable from

THE M-O VALVE CO. LTD

BROOK GREEN · HAMMERSMITH · LONDON W.6.

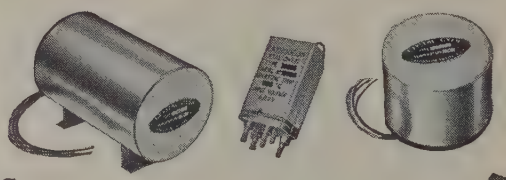
A subsidiary of the General Electric Co. Ltd

CATHODEON

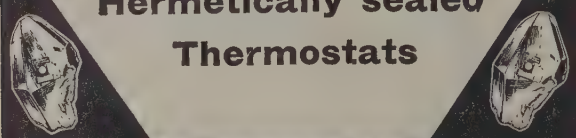
CRYSTAL OVENS

A complete range for fixed or mobile application achieving frequency stabilities

±0·0005% to ±0·000001%



Hermetically sealed Thermostats



CATHODEON CRYSTALS LIMITED

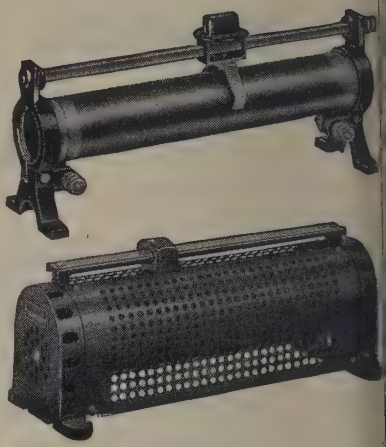
LINTON CAMBRIDGE ENGLAND TEL: LINTON 561

ZENITH

(REGD. TRADE-MARK

TUBULAR SLIDING RESISTANCES

Wound on vitreous enamelled steel tubes with oxidised wire having a negligible temperature co-efficient. Non-inductive windings on refractory formers are also available in many sizes.



Illustrated catalogue of all types free on request

The ZENITH ELECTRIC CO. Ltd.

ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN
LONDON, N.W.2

Telephone: WILlesden 6581–5 *Telegrams:* Voltaohm, Norphone, London

MANUFACTURERS OF ELECTRICAL EQUIPMENT
INCLUDING RADIO AND TELEVISION COMPONENTS

Publications of
THE INSTITUTION OF ELECTRICAL ENGINEERS

Journal of The Institution—Monthly

Proceedings of The Institution

PART A (Power Engineering)—Alternate Months
PART B (Electronic and Communication Engineering—including Radio Engineering)—Alternate Months
PART C (Institution Monographs)—In collected form twice a year

Special Issues

VOL. 94 (1947) PART IIA (Convention on Automatic Regulators and Servomechanisms
VOL. 94 (1947) PART IIIA (Convention on Radiocommunication)
VOL. 97 (1950) PART IA (Convention on Electric Railway Traction)
VOL. 99 (1952) PART IIIA (Convention on the British Contribution to Television)
VOL. 100 (1953) PART IIA (Symposium of Papers on Insulating Materials)
Heaviside Centenary Volume (1950)
Thermionic Valves: the First 50 years (1955)
VOL. 103 (1956) PART B Supplements 1–3 (Convention on Digital-Computer Techniques)
VOL. 103 (1956) PART A Supplement 1 (Convention on Electrical Equipment for Aircraft)
VOL. 104 (1957) PART B Supplement 4 (Symposium on the Transatlantic Telephone Cable)
VOL. 104 (1957) PART B Supplements 5–7 (Convention on Ferrites)
VOL. 105 (1958) PART B Supplement 8 (Symposium on Long-Distance Propagation above 30 Mc/s)
VOL. 105 (1958) PART B Supplement 9 (Convention on Radio Aids to Aeronautical and Marine Navigation)
VOL. 105 (1958) PART B Supplements 10–12 (International Convention on Microwave Valves)
VOL. 106 (1959) PART A Supplement 2 (Convention on Thermonuclear Processes)
VOL. 106 (1959) PART B Supplement 13 (Convention on Long-Distance Transmission by Waveguide)
VOL. 106 (1959) PART B Supplement 14 (Convention on Stereophonic Sound Recording, Reproduction and Broadcasting)
VOL. 106 (1959) PART C Supplement 1 (Position Control of Massive Objects)

Science Abstracts

Section A: Physics—Monthly.
Section B: Electrical Engineering—Monthly.
Cumulative Index.

Prices of the above publications on application to the Secretary of The Institution, Savoy Place, W.C.2.



RESISTORS

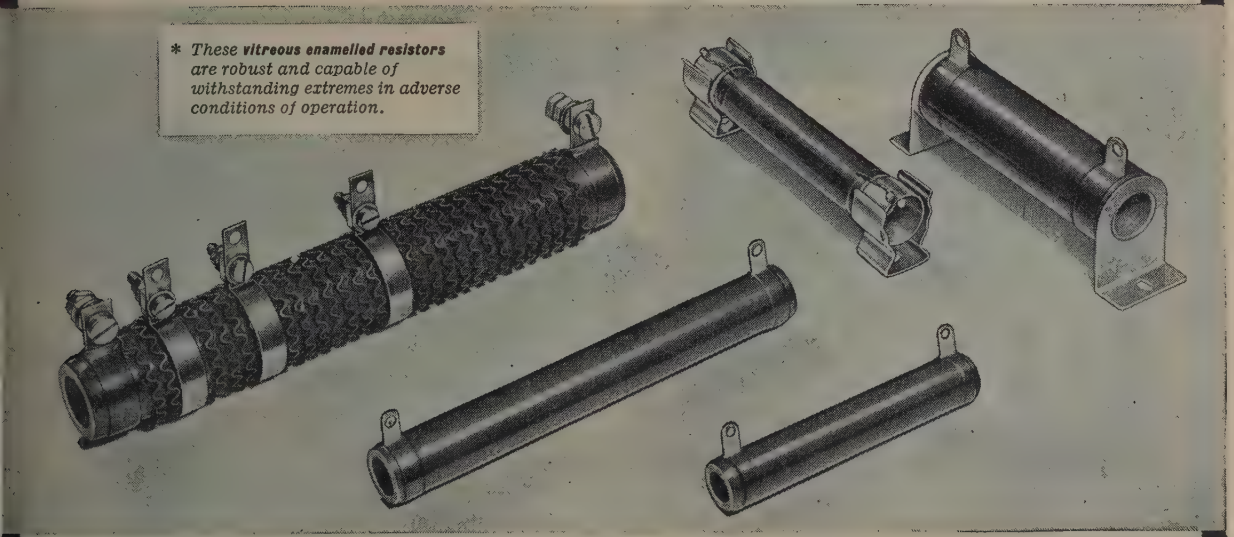
...from the largest...to the smallest

The smallest resistor engineered by Expamet Cressall is approximately of 4 watts rating. Above this, almost any size and type of resistor or rheostat can be supplied. Such a service is unique and backed by unrivalled experience. Technical advice that makes full use of this unrivalled experience is freely available.

EXPAMET and CRESSALL–go together all the way

* ***Wire wound resistors*** *on porcelain formers that can be pre-set by adjustable tapping bands to precise values.*

* *These* ***vitreous enamelled resistors*** *are robust and capable of withstanding extremes in adverse conditions of operation.*

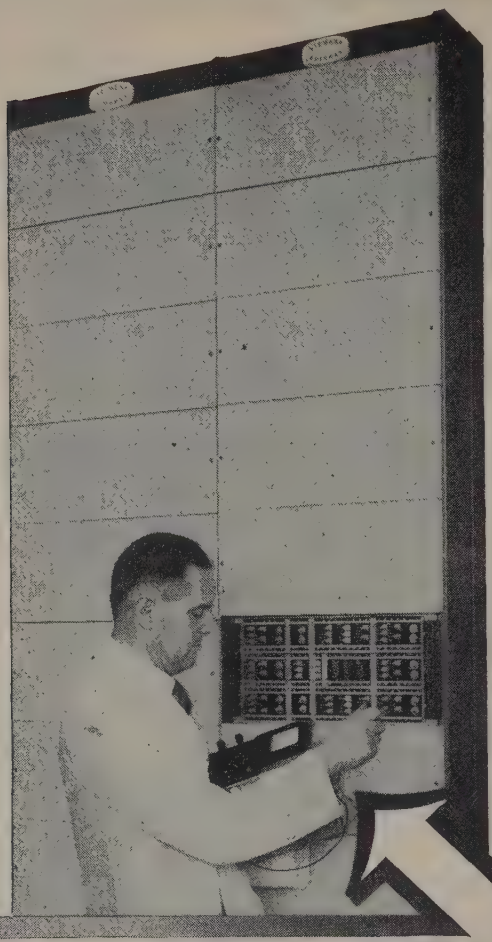


Expamet Cressall

The Electrical Division of The Expanded Metal Company Ltd.

LONDON OFFICE : Burwood House, Caxton Street, London, S.W.1. Telephone: ABBey 7766
WORKS : Stranton Works, West Hartlepool. Tel: Hartlepools 5531

The Cressall Manufacturing Company Ltd., Eclipse Works, Tower Street, Birmingham 19
Telephone: Aston Cross 2666

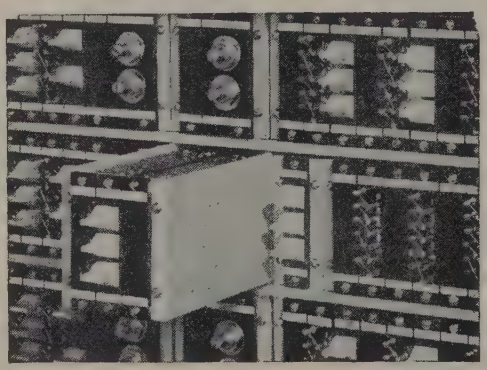


60 channels per bayside . . .

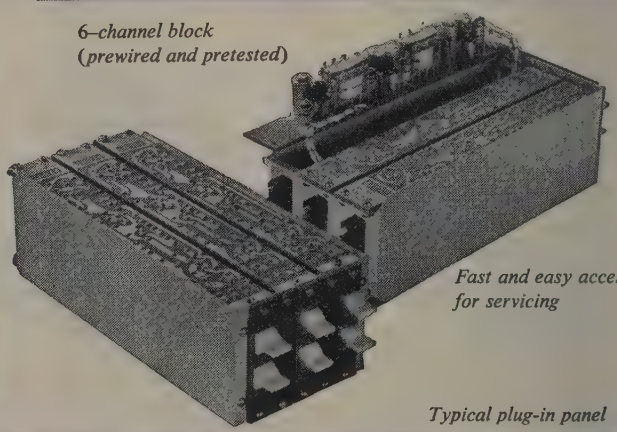
complete with carrier and power supplies

SIEMENS EDISWAN new equipment construction (E.C.3) offers this space saving advantage as well as these other features:

- 120 channel bay from 2 independent mountable baysides bolted back-to-back
- Complete with carrier and power supplies and inbuilt outband signalling
- Plug-in units
- Fully transistorised; designed and built to C.C.I.T.T. standards
- Complete and easy access to all components
- Uses standard 9 ft. by 20½ in. bays, allowing immediate incorporation into existing station arrangements
- Station cabling terminates at each 6-channel block



6–channel block (prewired and pretested)



Fast and easy access for servicing

Typical plug-in panel

vast experience  **in the science of telecommunications**

AEI

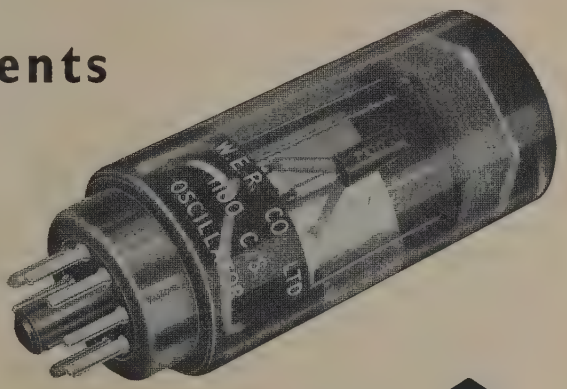
Telecommunications Division
155 Charing Cross Road London WC2 GERrard 8660
Associated Electrical Industries Limited

INDEX OF ADVERTISERS

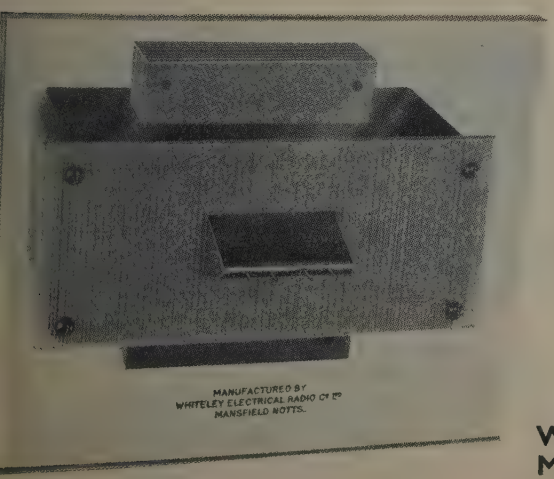
Firm	page
ola Products Ltd.	xxx
mec Ltd.	
W. G. Armstrong Whitworth Aircraft Ltd.	
ow Electric Switches Ltd.	vi
ociated Electrical Industries Ltd.	x, xiii, xxii, xxix, xxxiv
omatic Telephone & Electric Co. Ltd.	viii + ix
ing and Lee Ltd.	xvii
okhirst Igranic Ltd.	
le Makers Association	
hodeon Crystals Ltd.	xxxii
a (A.R.L.) Ltd.	
vhurst and Partner Ltd.	
iovan Electrical Co. Ltd.	
ilier Condenser Co. Ltd.	xix
o Electronics Ltd.	xxv
tro Dynamic Construction Co. Ltd.	iv
GA Products Ltd.	
lish Electric Co. Ltd.	xi
lish Electric Valve Co. Ltd.	xx
Resistor Co. Ltd.	i
.G. Industrial Corp. Ltd.	
ssons Telephones Ltd.	
anded Metal Co. Ltd.	xxxiii
ranti Ltd.	v
K. Fox Ltd.	
neral Electric Company Ltd. (M.O. Valves)	xxxi
neral Electric Company Ltd. (Semiconductors)	xvi
neral Electric Company Ltd. (Telecommunications)	xxvi + xxvii
rge Kent Ltd.	
ge Plugs Ltd.	
don Electric Wire Co. and Smiths Ltd.	xxx
ude Lyons Ltd.	xxiv
Marconi Instruments Ltd.	xii
Marconi Wireless Telegraph Ltd.	xiv + xxi
Metropolitan Plastics Ltd.	
M-O Valve Co. Ltd.	xxxi
Mullard Ltd. (Equipment)	iii
Mullard Ltd. (Components)	
Mullard Ltd. (Valves)	
Multicore Solders Ltd.	xxviii
Newmarket Transistors Ltd.	
Newton Bros. (Derby) Ltd.	xxx
Oliver Pell Control Ltd.	
Philips Electrical Ltd.	xxviii
Plannair Ltd.	xv
Plessey Co. Ltd. (Swindon)	vii + xviii
Racal Engineering Ltd.	
Rank Cintel Ltd.	xxiii
Richard Thomas & Baldwins Ltd.	
Salford Electrical Instruments Ltd.	
Savage Transformers Ltd.	
Semiconductors Ltd.	ii
Servomex Controls Ltd.	
Standard Telephones and Cables Ltd.	
Texas Instruments Ltd.	xxxvi
Telephone Manufacturing Co. Ltd.	
Ultra Electronics Ltd.	IBC
Whiteley Electrical Radio Co. Ltd.	xxxv
Zenith Electric Co. Ltd.	xxxii

WB electrical components

Our long experience in the design and manufacture of electrical components enables us to offer a wide range of products designed to meet with the many requirements of the electronic industry today.



This fixed frequency oscillator is constructed on a standard octal base and encapsulated in epoxy resin. Slight tuning either side of the fundamental frequency is provided by means of an adjustable core. Output: 10 mw into a 600 ohm. load. Frequency: As required within the range 700-2,000 c.p.s. sinusoidal.



On the left is an astatic wound A.F. screened line transformer with the windings encapsulated in epoxy resin. The insulation of the 'line' winding provides isolation against voltages of 30 KV RMS.

See us at the I.E.A. Exhibition where we will be pleased to make your aquaintance and discuss your individual requirements.

STAND NO. R 800

WHITELEY ELECTRICAL RADIO CO. LTD.
Mansfield, Notts. Telephone: Mansfield 1762/5

When you TRANSISTORISE

. . . LOOK TO TEXAS FOR

Experience pioneers of silicon transistors

Evaluation closely-defined, reliable parameters

Acceptance proven performance in world-wide service

. . . and you'll choose TEXAS

Texas are years ahead in silicon transistors, rectifiers and diodes . . . they have been in production in Great Britain since 1957. Texas components are to be found today in as widely differing circuitry as guided missiles, airborne electronics, radar and communications equipment, industrial instrumentation, process control systems and computers. They can answer *your* design problems *now*.

Joint Service Types
Many Texas transistors meet CV specifications. If you have a requirement for British Inter-Service types please write for full information.

Take a look at the TEXAS GROWN JUNCTION SILICON TRANSISTOR RANGE

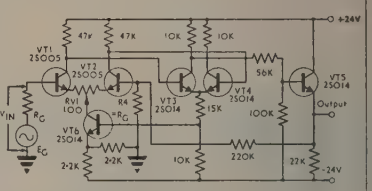


typified by the

2S014

D.C. AMPLIFIER

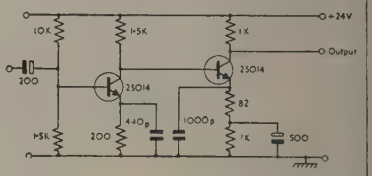
Voltage gain without feedback...	60dB
Voltage gain with feedback ...	40dB
Drift referred to input (temperature change 20°C to 70°C)—	
(1) using unselected transistors...	< 3mV
(2) with first pair matched for h_{FE}	∿ 1mV
Change of output voltage referred to input for simultaneous or independent changes of supply voltages by ±1 volt	< 4mV
Required source impedance ...	< 10 kΩ
Input impedance	25 kΩ
Common mode rejection ratio...	50,000



	2S002	2S003	2S004	2S014	2S005	Units
Guaranteed close beta spread	18-41	18-41	34-91	20-55	45-150	
High voltage	45	45	45	40	40	V
High cut-off frequency	7	15	8	20	30	Mc/s

VIDEO AMPLIFIER

Frequency response ...	10 c/s–3 Mc/s
Pulse response ...	0·2 μsec rise time
Power gain	30dB



Write for Data Sheets and Application Information on these and other TEXAS devices

TEXAS TELL YOU MORE You can design more accurately with Texas semiconductor devices, because they are completely evaluated in terms of closely defined and reliable parameters, and are consistent in quality and performance.

A.S.S.E. ELECTRICAL ENGINEERING EXHIBITION EARLS COURT OLYMPIA VISIT US ON STAND Y12

STOP PRESS!
ZENER VOLTAGE REGULATORS
AVAILABLE EX STOCK
8W. 15V to 150V.

TEXAS INSTRUMENTS
LIMITED
DALLAS ROAD BEDFORD ENGLAND
BEDFORD 68051 CABLES: TEXINLIM · BEDFORD

91/A4

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 107. PART B. No. 32. MARCH 1960

621.372.81.09

The Institution of Electrical Engineers
Paper No. 3132 E
Oct. 1959
©

SOME COMMENTS ON THE CLASSIFICATION OF WAVEGUIDE MODES

By A. E. KARBOWIAK, B.Sc.(Eng.), Ph.D., Associate Member.

(*The paper was first received 25th November, 1958, in revised form 10th February, and in final form 28th July, 1959. It was published in October, 1959, and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION *2nd November, 1959.*)

SUMMARY

The paper presents, as far as possible in non-mathematical language, the concepts of modes currently in use.

Some of the concepts are modified or generalized and a unified description is obtained. The term 'proper-mode' is reserved for a certain mathematical concept, which is usually concomitant with the concept of the eigen function, whereas the term 'quasi-mode' is reserved for a field configuration which can be supported by a physically realizable guiding structure.

In the description given, the radiation field which must always accompany a guided wave is viewed in a slightly novel light, and a bridge is constructed between the concept of modes guided in closed waveguides and those guided by open structures.

The paper concludes with a discussion on the existence of guided waves, and it is demonstrated that the existence or non-existence of a mode is entirely a matter of subjective opinion (implicit in the definitions). Nevertheless, it is shown that all quasi-modes are a physical reality and that they are useful in practical applications.

(1) INTRODUCTION

In an attempt to obtain a solution of a physical problem or to predict the behaviour of a physical system, one is led to combine the rigorous step involving a mathematical analysis with what could be called, philosophically, a weak link in the chain of the reasoning. A code for the translation of the physical attributes of the system into mathematical language is the weak link.

This translation can be carried out in an infinite variety of ways, and the mathematical methods employed subsequently in the analysis will differ accordingly, but provided that the code chosen (a set of definitions and concepts describing the physical phenomena) is correct, the physical interpretation of the results will, of necessity, be identical, though the labour involved in obtaining the results will probably be vastly different for different codes.

Thus, even if the inadmissible (logically or physically incorrect) methods of translation are excluded, one is still left with a number of possible approaches. The matter is further complicated by the fact that in most cases it is impossible to establish rigorously the correctness (or exactness) of a translation. Even if it were possible, a choice would still have to be made out of a number of concomitant concepts and definitions. At this point the guiding principle must be: which is the simplest and most useful approach?

For example, the concept of eigen functions is mathematically useful, frequently employed in various branches of theoretical physics, and its particularization in terms of modes is well known to those versed in waveguides. Nevertheless, in many practical applications it is useless and frequently does not lead to a clear physical picture capable of direct interpretation. The very question of the existence of such modes is difficult, if not impossible, to answer.

A question of practical interest requiring an answer is: can information be transmitted along a waveguide of a given structure in such a way that the power in the wave simply decays exponentially with the distance along the waveguide? More particularly, are there any waves whose field quantities vary with distance, z, as $\varepsilon^{-\gamma z}$ where γ is the complex propagation coefficient? The answer is yes, and modes of the type $f_\kappa(x_1, x_2)\varepsilon^{-\gamma_\kappa z}$ (hereinafter called the quasi-mode), where κ denotes the order of the mode and x_1, x_2 are the co-ordinates in the transverse plane, are useful in describing many physical fields, and a direct physical interpretation of such modes can be visualized, since every single one of them is a solution to the problem. In contrast, proper modes, unless combined in the right proportion, are not, on their own, solutions to physical problems.

Formally, in a source-free cylindrical region consisting of free space (or a perfect homogeneous and isotropic medium), the electromagnetic field is given by the solution of the wave equation

$$(\nabla^2 + k_0^2)\phi = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where ϕ is the wave function, $k_0 = 2\pi/\lambda_0$, the free-space wave number, and ∇^2 is the Laplacian operator (operating on co-ordinates x_1, x_2 and x_3).

The region is bounded by a surface (or surfaces) on which the electromagnetic field is required to assume some prescribed values.

Dr. Karbowiak is with Standard Telecommunication Laboratories, Ltd.

© 1960: The Institution of Electrical Engineers

Conventionally, ϕ is expanded in a series of some suitable set of orthogonal functions, and in principle the coefficients of this expansion are determinable by imposing the boundary condition on the surface. The problem is thus solved.

If the region is filled with a heterogeneous medium, the solution is obtained by splitting the regions into homogeneous sub-regions, obtaining a solution in each sub-region and matching the solutions across the dividing boundaries. Similarly, if the region is filled with an anisotropic medium, a rigorous solution is, in principle, again possible.

When dealing with practical problems, however, the satisfaction derived from being able to show that a rigorous solution is, in principle, obtainable is of little assistance, and in most cases the analytical and computational difficulties are insurmountable. What is even of greater importance to a physicist, and more so to an engineer, is that the rigorous approach is frequently devoid of the physical picture and does not lead to an intuitive understanding of the distribution of power flux, field intensity, lines of force, etc., i.e. all the tangible or semitangible attributes that can be imagined or measured.

It is, then, to help in forming the physical picture, and possibly to ease the solution of physical problems, that various concepts are introduced. These concepts, in general, are of utility in a limited class of problems and naturally cannot assist in the solution of all the problems.

Concepts are chiefly a matter of definition and cannot be proved or disproved. Also it is possible to start from two different concepts and arrive at the same quantitative result. Nevertheless, some concepts (operating in a given class) lead more quickly and more easily to the required result than do others. Again, from several co-existent definitions, one may be more general, or appear more appropriate, than the others, and yet be more cumbersome to apply. The question is not whether a definition or a concept is right or wrong, but how useful it is. This, however, is to a large extent a subjective measure and frequently has less to do with science than with engineering.

(2) SOME BASIC CONCEPTS

The variety of known waveguides can, according to their geometry, be broadly classified into uniform or cylindrical structures and non-uniform structures. They can be further sub-divided into closed and open guiding systems: the coaxial cable and hollow metal waveguides are examples of the former, while the twin-wire transmission line, the dielectric-rod waveguide and the single-wire transmission line are examples of the latter. When necessary, waveguides are further described by their cross-sectional geometry, e.g. linear, rectangular, circular etc.

The above are the geometrical attributes of the waveguide. The physical properties of the contained medium and the waveguide boundaries are its physical attributes. Thus, the medium filling the waveguide can be homogeneous or heterogeneous, isotropic or anisotropic, loss-free or perfect, etc., and similar adjectives could be applied in the description of the physical state of the waveguide surface.

(3) CLOSED WAVEGUIDES

(3.1) The Ideal Waveguide

If the bounding surface is perfectly conducting, the boundary condition to be imposed on the wave function, ϕ, can be of two types, leading to two distinct types of field:

Electric, resulting in E/TM fields:

$$[\phi]_s = 0 \quad (\phi \propto E_z) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Magnetic, resulting in H/TE fields:

$$\left[\frac{\partial \phi}{\partial n}\right]_s = 0 \quad (\phi \propto H_z) \quad . \quad . \quad . \quad . \quad (3)$$

If the equation to the boundary is of the type $x_1 =$ a constant, $x_2 =$ a constant, etc., in some suitable co-ordinate system, the eigen functions appropriate to that co-ordinate system are referred to as the proper-modes of the waveguide. Each mode on its own is then a solution to the problem and may be represented by

$$\phi(x_1, x_2, z) = \psi(x_1, x_2)\psi_3(z) \quad . \quad . \quad . \quad . \quad (4)$$

where

$$\psi_3(z) = \exp\left[j(\omega t \pm \beta z)\right] \quad . \quad . \quad . \quad . \quad (5)$$

$\beta =$ axial propagation coefficient, $z =$ axial co-ordinate and the function ψ is the solution to the two-dimensional wave equation*

$$\nabla_z^2\psi + h^2\psi = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$k_0^2 = h^2 + \beta^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

h is the cut-off coefficient.

In a separable system of co-ordinates† the function may be put in the form

$$\psi(x_1, x_2) = \psi_1(x_1) \, . \, \psi_2(x_2) \quad . \quad . \quad . \quad . \quad (8)$$

where ψ_1 and ψ_2 are called the circumferential and radial functions of circumferential (x_1) and radial (x_2) co-ordinates respectively.

The eigen values (wave numbers), h_{mn}, form a doubly infinite but discrete set of values determined by the boundary conditions. To each and every value h_{mn} there corresponds one (in a non-degenerate case, or several in a degenerate case) distinct eigen function, ψ_{mn}, whose nature is determined by the co-ordinate system chosen.

Evidently, if $h_{mn} < k$, then β is a pure real number and we are dealing either with a propagating mode, M_{mn}, or a standing wave (as in a resonator), M_{mnl}, where M stands for TM (or E) or TE (or H) as the case may be. For $h_{mn} > k$, β becomes an imaginary number and the mode an evanescent mode, denoted by $M_{mn\times}$.

In a similar manner, depending on the value of the wave-numbers, the function $\psi_1(x_1)$, apart from being a constant, could be oscillatory or evanescent. Analogous remarks would apply to $\psi_2(x_2)$.

Any mode that is evanescent in one of its co-ordinates is a form of surface wave. Thus $M_{m\times\bullet}$ represents an axial surface wave, whereas $M_{m\bullet\times}$ represents a radial variety of it. A surface wave need not be a travelling wave. For example, $TM_{0\times m}$ (or $E_{0\times m}$) in circular co-ordinates represents a mode in the surface-wave resonator resulting from resonating the $TM_{0\times\bullet}$ (or $E_{0\times\bullet}$) mode, a mode supported for example by a single-wire transmission line. Further, $TE_{01\times}$ (or $H_{01\times}$), in rectangular co-ordinates, can be looked upon as a resonant radial surface wave which results from resonating the $TE_{0\bullet\times}$ (or $H_{0\bullet\times}$) mode, or as a mode resulting from operating the conventional rectangular waveguide excited in $TE_{01\bullet}$ (or $H_{01\bullet}$) below cut-off.

By definition, the wave function of an axial surface wave

$$\phi = \psi_1(x_1)\psi_2(x_2)\varepsilon^{-j\beta z} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

must have the asymptotic form

$$\phi \simeq \psi_1(x_1)\varepsilon^{-ar}\varepsilon^{-j\beta z} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

* ∇_z denotes the two-dimensional Laplacian operator in the plane normal to the waveguide axis, z.

† This is a restriction made here for convenience, but it does not affect the general argument or our conclusions.

where r is the distance measured away from the surface (or surfaces) called the guiding surface, and a is the decay co-efficient (a positive real number).

The wave numbers are connected by

$$k_0^2 = \beta^2 - a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

For simplicity, the suffixes m, n, $\times$ or $\bullet$ associated with ψ, k, β, etc., will be omitted unless their retention is necessary to avoid ambiguity. Similarly, when there is no chance of confusion, the superscript (e) or (h) [as in $\psi^{(e)}$, $\psi^{(h)}$, or $h^{(h)}$, etc.] necessary to differentiate between quantities belonging to E and H modes respectively, will usually be omitted.

In a perfect waveguide any one of the proper modes could, given suitable terminal conditions, exist on its own. On the other hand, a number of proper-modes could propagate simultaneously along a waveguide and the proportion of each mode in the total field would be determined by the terminal conditions, but such a combination of proper-modes would not in turn be called a mode unless all the modes involved were degenerate. In the latter case, a useful meaning could be attached to such a field, since all the proper-modes involved have identical wave numbers, wavelengths, velocities, etc.; furthermore, the field distribution of a combination of degenerate modes is constant irrespective of the value of z and is determined entirely by the proportion of the modes involved.*

(3.2) Imperfect Homogeneous Boundaries

With imperfect (physical) waveguides, owing to radiation losses, losses in the boundary, etc., the cut-off coefficient, h, is no longer real but is usually a complex number, and consequently the propagation coefficient, β, also becomes complex.

In such waveguides, pure modes cannot exist and the distinction between propagating and evanescent modes is no longer clear, but with certain restrictions, provided the losses are reasonably small, the above classification into modes is still a useful one. More particularly, a proper-mode which could exist in a perfect waveguide becomes in an imperfect waveguide of necessity accompanied by other modes, so that the field is now representable by

$$M_0^{(I)} = M_0 + \sum_{i=1}^{N} \tau_i M_i + \tau_R R^{(I)} \quad . \quad . \quad . \quad (12)$$

where the finite summation extends over all relevant discrete modes and $R^{(I)}$ denotes the radiation field.

A useful quantity describing the purity of a mode, M, is the purity quotient, Q_M, defined by

$$Q_M = \left|\frac{P_M}{P_T}\right| = \frac{|P_M|}{|P_N + P_M|} \times 100\% \quad . \quad . \quad . \quad (13)$$

where P_M is the complex energy flow associated with the investigated mode, and P_N is the complex energy flow associated with all the remaining modes, including the radiation field.

It is now clear that, unless the purity quotient is high (all τ-values small), the whole concept of modes is of little assistance in forming a physical picture. Moreover, in such a case, the solution obtained for an imperfect waveguide by perturbation of the solution for a perfect waveguide is inaccurate and a rigorous solution with all its disadvantages must be sought.

The most common waveguide is formed by a metallic tube of constant cross-section. Strictly speaking, therefore, since the conductivity σ is finite, the energy of the guided wave cannot be confined entirely to the inside of the pipe, but will be partly guided within the metal wall and partly outside the pipe. Furthermore, to satisfy launching conditions, it is necessary to introduce the radiation field, and for this reason the total field cannot be a pure mode, but is describable by eqn. (12).

The solution to the general problem is of extreme complexity, but, since practical waveguides take the form of metallic pipes and the conductivity of metals is invariably a very large quantity (hence the skin depth is very small), it is permissible to neglect the field outside the pipe and replace the pipe by an imaginary boundary having a surface impedance Z_s ($|Z_s| \ll 1$). The solution can then be simply obtained by perturbation of the solution for a perfect waveguide.

Modes which can exist in a perfect waveguide become quasi-modes in an imperfect waveguide (and then only if such solution is permissible), but in most practical cases the quasi-modes differ very little from pure modes and the slight change in the phase velocity and the appearance of the attenuation coefficient are the more important features. It must nevertheless be remembered that the modes are not pure but are accompanied by other modes whose intensities are of the order of $|Z_s|$ as regards voltage.

(4) OPEN STRUCTURES

(4.1) General

In contrast to closed waveguides, where any field can be expanded in a discrete spectrum of proper-modes, the set of proper-modes appropriate for an open structure is not complete and the expansion must, in general, be supplemented by the radiation field.

This distinction between the two types of perfect guiding structures, coaxial and waveguide on the one hand, and twin-wire transmission and surface waveguides on the other, is rigorous and mathematically describable in strict terms and is of extreme importance in the classification of waves and the understanding of the whole subject of guided waves.

When dealing with imperfect waveguides, i.e. all physical ones, this distinction is not so obvious, and, in fact, rigorous mathematics will not help us. For, in the case of 'closed' waveguides as well as in the case of open ones, the description of the field in terms of proper-modes is not complete and the radiation field must be brought in.

However, even an approximate physical picture is better than none, and to that end, as with surface waves, we shall extend the concept of modes to cover the quasi-modes, which, as discussed above (if admissible in the solution of the problem), differ from the proper modes, and the appearance of the attenuation coefficient is the most important characteristic. The radiation field will be neglected and brought into the picture only where circumstances call for it.

(4.2) Surface Waves

It is evident from the form of the wave function of a surface wave [eqns. (9) and (10)] that for finite values of the decay coefficient, a, neither the wave function, ψ, nor its derivative $\partial\psi/\partial n$ can be made to disappear on the surface of the waveguide (x_2 = a constant). On the contrary, for example, with TM (or E) waves (where $\psi^e \propto E_z$), the E_z component over the surface of the waveguide must be of the order of Z_s, the surface impedance of the guiding surface, and since loss-free structures are being considered it is necessary that $Z_z = jX_z$, a pure imaginary.

Physical waveguides, however, are lossy, and therefore Z_s is of necessity complex ($Z_s = R_s + jX_s$), leading to perturbed wave numbers which are connected by

$$k_0^2 = -\gamma^2 - u^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where $\gamma = \alpha + j\beta$ = propagation coefficient
$u = a - jb$ = decay coefficient

and the mode becomes, by our definition, a quasi-mode.

* Examples: H_{01} (TE_{01}) and E_{11} (TM_{11}) in a circular waveguide, or E_{11} (TM_{11}) and H_{11} (TE_{11}) in a rectangular waveguide, etc.

Again, if the quasi-mode is to resemble the proper-mode, Z_s must be small and therefore a and b must be small ($|u| \ll k_0$). The attenuation, α, and the phase propagation coefficient, β, are then simply given by

$$\alpha = \frac{ab}{k_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

$$\beta = k_0(1 + \tfrac{1}{2}\epsilon) \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

with

$$\epsilon = \frac{a^2 - b^2}{k_0^2} + \left(\frac{ab}{k_0^2}\right)^2 \quad . \quad . \quad . \quad . \quad (17)$$

The radial propagation coefficient, u, which here replaces the cut-off coefficient, h, of closed waveguides, is a function of Z_s and the waveguide geometry. In a similar manner, as in the case of closed waveguides, h is a function of Z_s and the waveguide geometry.

Thus, for plane surface waves we simply have*

$$\left.\begin{aligned} a &= k_0 X_s \\ b &= k_0 R_s \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

so that

$$\alpha = k_0 R_s X_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

and

$$\beta = k_0[1 + \tfrac{1}{2}(X_s^2 - R_s^2 + X_s^2 R_s^2)] \quad . \quad . \quad . \quad (20)$$

Here, as with closed waveguides, the proper-modes exist only in a mathematical sense, but quasi-modes have their physical interpretation.

(5) LAUNCHING AND EXISTENCE OF MODES

(5.1) General

The concept of existence is entirely subjective, and in the mathematical as in the physical sense it cannot be derived from the definitions.

In an ideal or mathematical waveguide, a mode would be said to exist in a given field if, in the expansion into modes of the type shown in eqn. (12), the coefficient multiplying the mode sought was a finite quantity.

This definition, unfortunately, is not unique, because there is an infinite number of ways in which a field could be expanded and consequently a particular way of expansion must be chosen, i.e. in proper-modes.

With this definition, however, an immediate difficulty arises with open waveguides, since the proper-modes do not form a complete set. The difficulty can be circumvented by the introduction of the radiation field in addition to the proper-modes, and orthogonal with them. This is a possible and, mathematically, a permissible approach, yet philosophically it can lead to enormous difficulties, particularly if an extension to cover physical waveguides is attempted. Alternatively, we can resign ourselves to dealing with an incomplete set of functions and deal with problems where such a procedure is permissible; this procedure is preferred here.

(5.2) Possible Existence of Modes on Homogeneous and Quasi-homogeneous Structures

When dealing with closed waveguides, the purity quotient of a mode [eqn. (13)] is not only an indication of its existence but is also a quantitative measure of the proportion of the mode in the total field. Further, under quasi-static conditions the purity quotient must be a constant quantity independent of the axial co-ordinate, z, if a useful concept of a mode is to be preserved.

Every one of the proper-modes is, by definition, a solution to an ideal (a mathematical) waveguide, but not every quasi-mode corresponding to each of the proper modes is a physical possibility. Thus, for example, the TE_{11} (H_{11}) mode is a possible solution in a mathematical waveguide, but the quasi-TE_{11} (H_{11}) mode does not exist: on the other hand, the longitudinal section wave* ($EH)_{11}$ is possible in a mathematical and in a physical sense.

Such situations frequently arise; in fact, whenever an ideal waveguide capable of supporting simultaneously two or more degenerate modes is perturbed in one of its aspects, e.g. by slight bending, wave combinations which, by definition, were permissible in an ideal waveguide, may often not be admissible and may not be called quasi-modes.

The physical reason for this peculiarity is that, although in an ideal waveguide capable of supporting say two modes, M_1 and M_2, the combination solution $(M_1 + \tau M_2)$, where τ has any value, is also possible, yet, if the waveguide is perturbed in some uniform manner, the resulting waveguide will be capable of supporting the quasi-mode, $M_1 + \tau M_2$, only for certain definite values of τ determined by the nature of the perturbation. Such and only such a mode combination would have a purity quotient of nearly 100%, and this value would be independent of the value of the co-ordinate z despite the fact that the mode would suffer an attenuation.

Suppose, now, that the wall of the waveguide is corrugated or is endowed with some other suitable periodic property and that the period of the quasi-homogeneous structure is very small in comparison with the wavelength of the electromagnetic disturbance to be transmitted. It is then permissible to regard the waveguide as if it were homogeneous, of anisotropic surface impedance Z_s having a value determined by the state of the surface. The modes supported by such a waveguide are then very nearly the modes possible in an imperfect waveguide whose surface impedance has the value Z_s, save for a slight deterioration in the purity quotient. For a wave supported on a periodic structure is, by Floquet's theorem, of necessity accompanied by an infinite series (though discrete and related to the pitch of the structure) of space harmonics. If the pitch of the periodic structure is much smaller than the wavelength, the energy stored in the space harmonics is small in comparison with the power carried by the quasi-mode and consequently the purity quotient is high. As the pitch of the periodic structure increases, the purity quotient becomes smaller until eventually the wave structure bears no resemblance to the proper-mode from which it was derived.

In this classification no apology is made for applying the same approach to open structures and resonant lines. Thus, with surface-wave waveguides, if the wave exists when supported by a coated conductor, then it also must exist over any other guide (e.g. corrugated, a dielectric rod, etc.) having the same surface impedance as the coated conductor, since the nature of the wave is determined by the waveguide geometry and the value of Z_s.

Admittedly, we realize that, if a $TM_{0\times}$ ($E_{0\times}$) surface wave is guided on a corrugated waveguide, it cannot be in its pure form but must be accompanied by the space harmonics. Notwithstanding this, the purity quotient† will be high and the quasi-mode can be said to exist.

Furthermore, if a length of periodic surface-wave waveguide is short-circuited to form a resonator, the quasi-mode will still exist in the sense discussed. This again is permissible, despite the fact that if the system were loss-free it would radiate, the justification being that, if the pitch is small in comparison with

* Note that Z_s is surface impedance normalized with respect to Z_0 (377 ohms).

* A longitudinal section wave is a wave combination of E(TH) and H(TM) modes of the same order, such that one of the transverse field components is absent.

† With open structures, the purity quotient should be evaluated over an infinite area transverse to the axis of the waveguide. However, since the wave is evanescent over the wavefront, it is permissible and more convenient to evaluate the purity quotient over a finite area extending sufficiently far to include, say, 99·9% of the power in the surface wave; the exact choice of the size of this area is not critical.

the wavelength, the power lost by radiation need not be greater than the ohmic loss in the structure, a small quantity.

Examples of other quasi-homogeneous structures are given by waveguides filled with laminar media where the laminar thickness is much smaller than the wavelength.

(5.3) Launching of Modes

Although a given structure may be capable of supporting a mode (a proper-mode in an ideal waveguide and a quasi-mode in a physical waveguide), the question of existence still remains unanswered. Clearly, if a field approximating to that of the investigated mode could be created over the entire cross-section of a suitable waveguide using a suitable antenna system, then the problem of existence would be resolved. The question now arising is: how well must the antenna field resemble the mode in question, and is it physically possible to construct a suitable antenna? In the case of closed ideal waveguides the question can be answered on a rigorous mathematical basis, and it transpires that all proper-modes can be launched, even from a point source, provided that it is placed and polarized suitably (this can be decided from geometrical considerations). The determination for a given antenna of the purity quotient and the proportion of each of the propagating modes in the total field may, however, be a matter of lengthy mathematical manipulation. In principle, it can be done.

With imperfect waveguides, however, the mathematical complexity becomes immense, and no solution is yet known. Moreover, it is very unlikely that a rigorous mathematical solution to the problem of launching waves in imperfect guides will be forthcoming in the near future.

The mathematical aspects of this problem can be sketched briefly as follows. The solution for the field launched from an antenna is made up from integrals, each evaluated in the complex plane of the propagation coefficient, γ. In this complex plane, there are distributed branch points and a large number of poles. Each of the integrals involved is a sum of branch-cut integrals (representing the radiation field) and the pole residues (representing the possible quasi-modes). Thus, if the coefficients of the expansion could be determined, a useful meaning could be attached to a theoretical value of the purity quotient and the question of existence settled in a philosophically satisfactory manner. But the solution would normally be obtained by developing the closed-form expression into an asymptotic series, and asymptotic series are not unique in the sense that any function proportional to $\varepsilon^{-\alpha r}$ (α real, r = distance from the antenna) could be added to or subtracted from the solution. Since all quasi-modes contain such an exponential factor, the method of asymptotic series, though useful for determining the radiation field, is of little assistance in a rigorous determination of the purity quotient.

Similar remarks would apply to an open guiding structure, but here the proportion of the radiation field in the total field is considerably larger. The difference is only quantitative and not qualitative.

The difference between an open and a closed waveguide becomes apparent, however, in the limit, when dealing with perfect waveguides. A perfect closed waveguide excited at a frequency above the cut-off of the lowest-order mode, but below the cut-off of the next higher-order mode, will transmit the lowest-order mode with a 100% purity quotient. On the other hand, a perfect surface-wave waveguide cannot carry a 100% pure surface-wave mode, because of the contamination by the radiation field.

Physical waveguides are, however, lossy and therefore all waveguides, whether closed or open, support a quasi-mode contaminated by the radiation field. Consequently, if launching

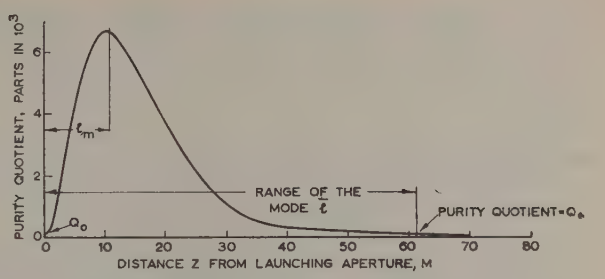


Fig. 1.—The purity quotient as a function of distance from the launching aperture, in the presence of a radiation field.

It is assumed that power in the guided mode is proportional to $\varepsilon^{-\alpha z}$, and power in the radiation field, to z^{-2}.

Constants of the guiding structure:
Q_0 = launching purity quotient = 10.
α = 0·1 neper/m.

conditions are to be considered, the purity quotient (Fig. 1) becomes a function of distance, since the radiation field is proportional to z^{-2} and the guided mode is proportional to $\varepsilon^{-\alpha z}$, and the conventional concept of modes is of little significance. For this reason, when discussing waveguide modes, the radiation field will be neglected, and only when special circumstances call for it will it be brought back into the picture.

(6) A NOTE ON CLASSIFICATION OF SURFACE WAVES

Conventional closed waveguides have now been studied for quite a long time, and, if the conventional nomenclature is qualified by the remarks in the previous Sections, a satisfactory though approximate classification results.

To arrive at a satisfactory definition and classification of surface waves it is convenient (Section 5.3) to neglect the radiation field (i.e. to assume that the surface wave can be launched, and only if necessary to investigate later if this is indeed possible) and start with the definition of proper-modes on loss-free structures. The definition can subsequently be extended to cover the quasi-modes.

Basically, as with the modes in the closed waveguides, surface waves can be subdivided into TM (or E) and TE (or H) modes as well as the hybrid waves (EH modes), and all waves can be classified according to the geometry of the co-ordinate system employed in relation to the shape of the guiding surface.

The basic characteristic in common with all surface waves is that they are evanescent over the wavefront; they are therefore of the form $M_{m\times\bullet}$. The $TM_{0\times\bullet}$ (or $E_{0\times\bullet}$) is the most common type of surface wave and its wave function has the asymptotic form

$$\phi \simeq \psi_1(x_1)\varepsilon^{-ar}\varepsilon^{-j\beta z} = E_z \quad . \quad . \quad . \quad . \quad (17)$$

The $TE_{0\times\bullet}$ (or $H_{0\times\bullet}$) mode is also possible on cylindrical structures, but the $TM_{0\times\bullet}$ (or $E_{0\times\bullet}$) is the principal mode, having no low-frequency cut-off.

All the remaining higher-order modes can usually exist only in the hybrid form $(EH)_{m\times\bullet}$.

A surface wave can only be supported by reactive surfaces, $Z_s = jX_s$ (X_s must be a finite positive number), and the relation between the coefficients are given by eqns. (11)–(16). The quantity Z_s must be homogeneous or quasi-homogeneous. Heterogeneous surfaces cannot support waves as defined here, but terms of the type 'leaky surface waves' and 'quasi-trapped surface waves' may be useful descriptions.

Examples.—(*a*) A wave guided over a plane inductive homogeneous or quasi-homogeneous sheet is a plane surface wave (a proper mode), the $TM_{0\times\bullet}$ (or $E_{0\times\bullet}$) mode. Irrespective of

whether the guide is a dielectric-coated surface, a corrugated surface (quasi-homogeneous), a dielectric slab, etc., the mode supported would be identical in all cases, and the extent of the field would also be the same provided that the Z_s values were identical.

If the structure is lossy, a quasi-mode results. The particular mode supported on a surface for which $R_s \geqslant X_s$ is called a Zenneck wave, but whether the guiding structure is a conducting half-space, or lossy dielectric coating over a conducting plane or any other structure is entirely immaterial, since the values of the propagation coefficients that determine the nature of the Zenneck wave are entirely determined by the magnitude and the proportion of R_s and X_s, irrespective of what the actual physical structure of the waveguide happens to be.

(*b*) A wave supported by a circular cylindrical inductive surface is a circular axial surface wave, the $TM_{0\times\bullet}$ (or $E_{0\times\bullet}$) mode.

Remarks similar to those in (*a*) about the surface impedance are also applicable here. The quasi-mode supported on a surface for which $R_s = X_s$ (a metallic surface) is called the Sommerfeld wave. A wave supported on a dielectric-coated waveguide is called the Harms wave, an unnecessary distinction since the wave is no different from that supported on a dielectric rod, a corrugated rod, etc., having the same value of Z_s.

(*c*) A wave supported on a helical waveguide is a hybrid wave $(EH)_{0\times\bullet}$.

(*d*) A 'dipole mode' supported on a dielectric rod is an example of another hybrid mode, an $(EH)_{1\times\bullet}$ mode.

The radial surface waves have the general form $M_{m\bullet\times}$, indicating that the wave is propagating in the radial direction and is evanescent in the axial direction.

Further terms could be coined to order, but the use of the clumsy nomenclature can be avoided if the index notation is accepted and adhered to, as is customary with closed waveguides.

(7) CONCLUSION

Various concepts of modes have been discussed. Their usefulness in describing electromagnetic waves in physical and idealized waveguides has been analysed and some novel concepts have been introduced.

It transpires that there is an infinite variety of ways in which modes could be defined, but the more important findings are as follows:

(*a*) Proper-modes are modes appropriate for a perfect waveguide, which is a mathematical concept of a loss-free waveguide. Such modes do not exist, but they are useful in some applications involving short lengths of guide. With closed waveguides they form a complete set, but with open structures the infinite set of proper modes must be supplemented by a continuous spectrum called the radiation field.

(*b*) Quasi-modes are fields whose intensities decrease with the axial distance z as $\exp -(\gamma z)$, where γ is the propagation coefficient of the mode. They are extremely useful in descriptions of physical fields and many physical fields approximate to quasi-modes very closely indeed, but strictly speaking quasi-modes do not exist in a pure form and the set of modes is incomplete and should be supplemented by the radiation field, whether the waveguide be open or closed.

(*c*) With physically realizable waveguides there is associated a field configuration that changes with the axial distance z. With many physical waveguides, however, over a certain finite range of the axial distance of interest in applications, the field resembles very closely a quasi-mode, and hence the importance of the concept of the quasi-mode. Such quasi-modes are said to exist. Others, on the other hand, cannot be obtained in a sufficiently pure form over any finite range, and such modes must be classified as physically non-existent in the sense that the amplitude of the mode in the total field is insignificantly small.

The radiation field in waveguide problems has been put in its proper perspective and discussed at length. It transpires that with all physically realizable waveguides, the radiation field dominates at large distances from the source, but over moderate distances quasi-modes retain their significance.

Surface waves have been discussed in some detail, particularly because many of the less obvious characteristics of guided waves are brought into better light through the study of these waves.

The discussion has been concluded by a classification of waves in terms of the modified and unified index notation, which helps to put the closed and open guiding systems on a more harmonious basis.

(8) ACKNOWLEDGMENT

Acknowledgment is made to Standard Telecommunication Laboratories, Ltd., for permission to publish the paper.

(9) BIBLIOGRAPHY

(1) STRATTON, J. A.: 'Electromagnetic Theory' (McGraw-Hill, 1941).

(2) SCHELKUNOFF, S. A.: 'Electromagnetic Waves' (Van Nostrand), p. 943.

(3) HUXLEY, L. G. H.: 'A Survey of Principles and Practice of Waveguides' (Cambridge University Press, 1947).

(4) Proceedings of the Symposium on Modern Advances in Microwave Techniques, New York, 1954.

(5) KARBOWIAK, A. E.: 'Theory of Imperfect Waveguides: the Effect of Wall Impedance', *Proceedings I.E.E.*, Paper No. 1841 R, September, 1955 (**102 B**, p. 698).

(6) KARBOWIAK, A. E.: 'Waveguide Characteristics', *Electronic and Radio Engineer*, 1957, **34**, p. 379.

(7) BARLOW, H. M., and CULLEN, A. L.: 'Surface Waves', *Proceedings I.E.E.*, Paper No. 1482 R, April, 1953 (**100**, Part III, p. 329).

(8) TAI, C. T.: 'The Effect of a Grounded Slab on the Radiation from a Line Source', *Journal of Applied Physics*, 1951, **22**, p. 405.

(9) WHITMER, R. M.: 'Radiation from a Dielectric Wave Guide', *ibid.*, 1952, **23**, p. 949.

(10) WHITMER, R. M.: 'Fields in Nonmetallic Waveguides', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1105.

(11) GOUBAU, G.: 'On the Excitation of Surface Waves', *ibid.*, 1952, **40**, p. 865.

(12) CULLEN, A. L.: 'The Excitation of Plane Surface Waves', *Proceedings I.E.E.*, Monograph No. 93 R, February, 1954 (**101**, Part IV, p. 225).

SOME COMMENTS ON QUASI-OPTICAL METHODS AT MILLIMETRE WAVELENGTHS

By L. Lewin, Associate Member.

(*Summary of a contribution which was read before the* Electronics and Communications Section, 2*nd November*, 1959)

At wavelengths less than about 1 cm, and particularly at the shorter millimetric wavelengths, the usual rectangular waveguide methods of transmission and component design become more and more impracticable, owing to the very small dimensions and increasing attenuation. For example, at a wavelength of 4 mm the recommended size of guide aperture is 0·061 in × 0·122 in and in copper this leads to attenuation of the order of 3 dB/m. At wavelengths of 8 cm many components require designs involving tolerances of the order of a few mils, and it is obvious that it is impracticable to scale these designs down by a factor of 20 or more. If the lower end of the microwave region is to be fully exploited, a transmission medium must be devised of sufficiently low loss and adequate dimensions for practical use.

The name 'quasi-optical' has been given to the field in which energy produced in a small waveguide is radiated and then acted upon by such 'optical' elements as prisms, lenses, gratings, etc. Strictly speaking, such methods are not so much quasi-optical as genuinely optical techniques utilized in an unsuitable region where, because of the small size of apertures in terms of wavelength, diffraction effects predominate. Many difficulties have been met from this cause, and what is sought here is a technique which is truly intermediate between the current waveguide and orthodox optical methods, and which lends itself readily to component design and manufacture.

Six categories of transmission are evaluated, determined by whether, in each of the two directions defining a cross-section perpendicular to the direction of propagation, one, several, or an infinite number of modes can propagate. Of these six categories the over-moded square or circular waveguide operated in the dominant mode is preferred.

Some suggestions are made for a number of components in this medium. The generation of spurious higher-order modes is a problem. For a square guide whose side is only 4λ long, a right-angle corner with a 45° mirror at the corner had a measured loss to the main mode of about ½ dB. Methods of accurately measuring and controlling spurious modes have yet to be developed.

Mr. Lewin is with Standard Telecommunication Laboratories, Ltd.

DISCUSSION ON THE ABOVE PAPER AND CONTRIBUTION BEFORE THE ELECTRONICS AND COMMUNICATIONS SECTION, 2ND NOVEMBER, 1959

Professor H. E. M. Barlow: Dr. Karbowiak has made the problem of distinguishing different waveguide modes rather more difficult than it need be, and his insistence upon the term 'quasi' does not help. Some of the wave configurations he includes as so-called quasi-modes are, in fact, proper-modes, and I cannot see why in those cases he adopts the nomenclature suggested. For example, in a circular-section guide we can have any of the H_{0n} or E_{0n} families of wave modes and they are suitably described as proper-modes because each can exist in a guide of finite wall impedance without any other modes being present at the same time. Moreover, surface waves supported by lossy surfaces which are straight in the direction of propagation are also proper-modes. These are all cases in which we can obtain a modal solution to the Maxwell equations complete in itself and satisfying the boundary conditions, so that there seems no point in introducing the term quasi-mode.

There are, of course, other cases in which we do not find this simplicity, e.g. modes in rectangular waveguides or those in circular waveguides other than the H_{0n} and E_{0n} families, where a finite surface impedance requires that additional modes should be present at the same time, causing contamination and producing a resultant field pattern which may be properly described as a quasi-mode. But I feel that we should reserve this term quasi-mode for what can be legitimately regarded as perturbations of the corresponding proper-modes, and in this way we would meet most of the difficulties to which attention has been drawn. For example, the well-known TEM mode has, in fact, no real existence, because any actual guiding structure must have a finite surface impedance, introducing a longitudinal component of the electric field, but we do not as a rule call this a quasi-TEM mode. We accept the fact that our definition is based on an ideal system and I feel sure this is the right approach, enabling us to deal with the matter in the simplest possible way.

Having defined our modes in terms of idealistic conditions, we can then recognize the practical situation as introducing certain modifying influences resulting in the creation in some cases of impure or quasi-modes, still having the general character of the basic proper-mode. These proper-modes really represent a mathematical concept, and the difficulty lies in reconciling this concept with the physical situation. In accepting the mathematical definition of a proper-mode we are doing the best we can in the circumstances, and it does little harm to acknowledge a slight departure from the ideal conditions to encompass the requirements of a practical system, provided that we recognize and understand what is happening. Dr. Karbowiak would thus do well to review some of his mode classifications and take into account more specifically our traditional approach to this problem.

I am not clear on what ground Mr. Lewin chooses the square- and circular-section guides in the dominant modes. These guides obviously have many advantages, but they also have their disadvantages, and I should like to know more about the reasons for his conclusion. There seems quite a good case for the single-wire transmission line in certain applications, and the reluctance to adopt it as a practical waveguide is unjustifiable. Calculation and experiment show that a single bare copper wire of about 36 s.w.g. supporting a 35 Gc/s surface wave has an attenuation of 1 dB/m in a straight length. At 70 Gc/s the attenuation rises to about 1½ dB/m, but this performance compares very favourably with some of the other waveguides to which Mr. Lewin refers. It is true that the surface wave has a tendency to radiate at a bend, but much can be done to avoid this by coating the wire with a thin layer of dielectric. Even with a bare copper wire the field decays rapidly outside the surface, falling to about 1/50th of its surface value at a radius of ½ cm. A very large proportion of the power is therefore transmitted close to the surface of the conductor. Experiment shows that we can bend a dielectric-coated wire without adding significantly to the attenuation or altering the electrical length. The waveguide itself could not be simpler, and a surface wave can be readily launched over it by a horn extending from a short length of coaxial line coupled to the rectangular waveguide supply. I therefore suggest that the single-wire transmission line is worthy of careful consideration for some millimetre-wave applications.

Dr. J. Brown: Any new concept or definition should either simplify the existing treatment of a subject or clarify some point of difficulty; I see no essential difference between Dr. Karbowiak's proper-modes and quasi-modes, and thus consider that the introduction of this subdivision is an unnecessary complication which does nothing to further our understanding of the subject. He defines a proper-mode as a field which satisfies Maxwell's equations and the boundary conditions and which has a configuration of the type $f(x, y)\varepsilon^{-\gamma z}$ where x and y are transverse, and z axial, co-ordinates. It is unnecessary to restrict this definition to loss-free structures, and many examples exist of modes, defined in this way, in lossy structures. A simple case is a circular waveguide with walls of finite conductivity,* and a set of modes which satisfies the above definition exists for this structure.

I cannot find a corresponding explicit definition for the quasi-mode in the paper. This concept is introduced by eqn. (12), which states that the field can be represented by a sum of quasi-modes and a radiation field. An extra complication is introduced in that attention is shifted from the question of whether modes in the sense of the above paragraph can exist to the representation of the field excited by a specific source. Will Dr. Karbowiak define a quasi-mode explicitly?

If the concepts of proper-modes and quasi-modes are both essential, there must be significant differences between the two. Three possible differences mentioned in the paper are

(*a*) There is a clear-cut distinction between propagation and evanescence for proper-modes but not for quasi-modes.
(*b*) The transverse field functions for proper-modes form a complete set of functions, but those for quasi-modes do not.
(*c*) Proper-modes can exist as single entities but quasi-modes cannot.

These three distinctions do not, in fact, seem to be established because

(*a*) A radially expanding cylindrical wave, such as could exist in a sectoral horn, satisfies the basic definition of a proper-mode. At large distances from the throat of the horn this wave has a propagating character, but near the throat it may have an evanescent character. There is no sharp distinction between the propagating and evanescent regions.
(*b*) A surface wave supported by a loss-free surface is admitted as a proper-mode, but may be the only modal solution. There is certainly never a complete set of modes in such a situation.
(*c*) This point can be resolved only if a precise definition of the quasi-mode is given. Confusion can arise by arguing from the special case of a loss-free guide to the more general one of a lossy guide. It is impossible to make deductions about a general case from a special one, which is what is being attempted in the last paragraph of Section 3.2.

Some very valuable comments on the concept of modes in non-conventional waveguides are contained in the report of the Brooklyn Polytechnic Symposium.*

Mr. R. A. Waldron: The mathematical technique in investigating the properties of an oscillatory system is to start with the wave equation and solve it. For a waveguide the solutions are the electromagnetic field components in terms of certain arbitrary constants. By applying the boundary conditions, a set of linear homogeneous equations in the arbitrary constants is obtained, there being one more equation than there are arbitrary constants. The condition for the consistency of these equations is that the determinant of the coefficients of the arbitrary constants shall vanish. This condition gives the characteristic equation, which can be solved for the eigenvalues of the system. For a waveguide the eigenvalue is the phase coefficient, β. Knowing β, the arbitrary constants can be solved for, and hence the field components determined. Corresponding to each eigenvalue there will be a distinct set of field components. It is these sets of field components that define what is meant by a mode.

This technique is always possible in principle, although in practice it is often difficult to carry out. Normally, the physical system being dealt with is complicated, and we seek a model which is physically similar but mathematically simpler. Thus the usual waveguide consisting of a metal tube containing air can be treated mathematically as a perfectly conducting tube containing a vacuum. But this model cannot be used for a guide whose walls are not metallic, or which contains something other than a vacuum.

It is not a question of what is meant by a mode; there is no difficulty about this. The crux of the matter is the choice of an appropriate model.

Mr. L. Lewin: Dr. Brown says that Dr. Karbowiak's distinction in relation to the phase coefficient—that it is purely real or purely imaginary in the proper-mode and mixed in the quasi-mode—is not acceptable, because it breaks down in a horn, where we have an eigenfunction with both characteristics of behaviour. As I understand it, the whole of the author's discussion is limited to the rectilinear case, so that this eigenfunction does not enter the picture and the distinction holds. Within this limited picture the author says that if there is a wall with impedance Z_s the quasi-mode concept must be used and the perturbed features are small if Z_s is small. If Z_s is complex, which means that the wall is lossy, the distinction can be drawn that the phase coefficients are purely real or purely imaginary in the proper-mode case and mixed in the other; but if Z_s is purely reactive, does the author think that it is a proper-mode or a quasi-mode in relation to the definitions which he uses?

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Dr. A. E. Karbowiak (*in reply*): The paper is largely non-mathematical, but there is evidence in support of the statements made.† Definitions, provided that they are logically consistent, can be chosen arbitrarily and the results depend on the choice. The speakers are mostly concerned with the correctness of definitions and I must stress that logically this is beyond discussion. If Prof. Barlow wishes to call E_{0n} and H_{0n} as well as TEM modes in physical waveguides proper-modes, and Dr. Brown calls $f(x, y)\varepsilon^{-\gamma z}$ a proper-mode of a physical waveguide, I cannot prove or disprove it, because these are merely suggested definitions.

What is important, however, is the usefulness of a concept: this is the topic interwoven in the whole of the paper, and appreciated rightly by Mr. Waldron in his last sentence. The models mainly advocated by the speakers are those commonly exploited, but these are applicable only to a small class of problems, whereas the discussion presented in the paper is of wide application.

The examples given by Prof. Barlow and Dr. Brown illustrate very simple and idealized cases, but even there I do not see how they can say that there is not much wrong with their model considering that during the last few decades hundreds of papers have been published on the existence of surface waves and yet the opinions are divided. Furthermore, there have been numerous international discussions on the meaning of modes in

* STRATTON, J. A.: 'Electromagnetic Theory' (McGraw-Hill, 1941), p. 524.
† See, for example, KARBOWIAK, A. E.: 'Radiation and Guided Waves', U.R.S.I. International Symposium on Electromagnetic Theory, June, 1959, Toronto, Canada.

* 'Modern Advances in Microwave Techniques' (Polytechnic Institute of Brooklyn, 1955), p. 481.

other than simple waveguides (see, for example, Reference 4) with little agreement. Surely the physical picture is at fault, and the case presented by the speakers is very one-sided.

New concepts and definitions are frequently introduced because they have a wider range of validity and usefulness; this is the case with quasi-modes. A quasi-mode forms part of a field which can exist in a physical waveguide and has the form $f(x, y)\varepsilon^{-\gamma z}$. It differs from a proper-mode in a number of respects. Proper-modes are eigen functions (solutions to the homogeneous wave equation) of a perfectly conducting waveguide and form an infinite complete set. This is a mathematical concept describable in rigorous terms. When applied to a physical waveguide the presentation becomes, invariably, an approximation. Suitably perturbed proper-modes can be interpreted again as proper-modes of a waveguide with surface-impedance boundaries: this is a permissible presentation. These modes are of the type $f(x, y)\varepsilon^{-\gamma z}$ and form an infinite set which is complete in a closed waveguide. Furthermore, these modes may be quasi-modes of a physical waveguide, in which case they do not form a complete set, or even an infinite set [e.g. eqn. (12)]. They are called quasi-modes because they need not be solutions to the wave equation and, in contrast to proper-modes, they need not, on their own, satisfy all boundary conditions (either of surface-impedance type or radiation condition at infinity), but, in finite regions they are frequently solutions (almost) to the inhomogeneous wave equation applicable under physical conditions.

The distinction is sometimes subtle but nevertheless essential, and most of Dr. Brown's difficulties arise because, in the simple cases he is considering, quasi-modes differ very little from proper-modes, but a more detailed consideration does bring out the differences. His differences (*a*), (*b*) and (*c*) are, on their own, not really essential, and (*a*) is incorrect.

The E_{0n} and H_{0n} modes of a physical waveguide, to which Professor Barlow refers, cannot be regarded as proper-modes for the reasons given above. They can exist as proper-modes in a waveguide with surface-impedance boundaries but hardly in a pure form, except to first order of quantities.

Dr. Brown's insistence on rigour does not ease the situation, since no solution to a physical problem can be really rigorous. Solutions to a mathematically describable model can certainly be rigorous, but his model is only an approximation to a physical situation. The Stratton analysis is really a disguised surface-impedance approach and the modes are therefore proper- and not quasi-modes. In fact, the case of a waveguide with metallic walls has not been treated rigorously until recently (*loc. cit.*). Such analysis reveals the finer points that I have been discussing. Although these points may be unimportant in the case of metallic waveguides (tubes), most of the confusion in the past concerning surface waves and other more complicated structures is due to the failure to appreciate what happened to be 'unimportant' in metallic waveguides.

The solution which Mr. Waldron advocates leads through a homogeneous wave equation to proper-modes, but the analysis gives no clue to the physical existence of such modes; in fact, many of the modes are inadmissible. A better approximation is a solution to the inhomogeneous wave equation, but this leads to representation in terms of a contour integral which invariably is not conducive to thinking in terms of correct physical models.

Quasi-modes, on the other hand, are linked with physical existence. Sometimes a given field structure may be a proper-mode of one kind of waveguide and a quasi-mode of another, or it may be inadmissible. For example, in a waveguide with surface-impedance boundaries an E_{0n} mode is a proper-mode, but in a metallic waveguide it is a quasi-mode in some region if n is not too large and is not admissible if n is very large. In a similar way the Zenneck wave, as we know it, can be regarded as proper-mode in the sense that it is a solution to the homogeneous wave equation, and yet as a quasi-mode it has no physical significance.

The discussion given in the paper, although illustrated by uniform waveguides, is for the most part also applicable to non-uniform waveguides.

All physical waveguides are lossy and, therefore, strictly speaking can support only quasi-modes; reactive waveguides, on the other hand, could be regarded under suitable conditions as supporting proper-modes.

Mr. L. Lewin (*in reply*): At 35 Gc/s the attenuation of standard (size 22) copper waveguide is less than the figure quoted by Prof. Barlow for the surface-wave mode on a straight wire and it gives no special trouble at bends. Nevertheless, its attenuation is excessive for certain purposes (e.g. the loss accompanying high-Q-factor filter elements) whilst the tolerances on the dimensions of some components may be too tight to be practical. It is in order to overcome these two limitations that the over-moded square or circular guide has been proposed.

I cannot see in what way the surface-wave transmission line could overcome these defects, and it seems to be an even more unsuitable medium than rectangular guide for making high-quality components. This is not to say that it has not a certain field of usefulness but only that it, and also standard waveguide, is not practical for some high-quality purposes. Whether the over-moded guide will be adequate has yet to be demonstrated.

621.317.79 : 621.382

The Institution of Electrical Engineers
Paper No. 2962 M
June 1959
©

AN ANALOGUE ELECTRONIC MULTIPLIER USING TRANSISTORS AS SQUARE-WAVE MODULATORS

By P. GLEGHORN, B.Sc.(Eng.), Graduate.

(*The paper was first received 7th November*, 1958, *and in revised form 12th February*, 1959. *It was published in June*, 1959, *and was read before the* MEASUREMENT AND CONTROL SECTION *3rd November*, 1959.)

SUMMARY

The need for a relatively high-speed analogue multiplier is discussed and the types of multiplier likely to meet this need at reasonable cost are investigated. It is concluded that the modulated square-wave type of multiplier using transistor modulators will be most suitable.

The extension of the scope of the multiplier to operate with inputs of either sign is discussed. The causes of the various errors in the multiplier are examined and an overall accuracy of within $\pm 2\%$ is quoted. An explanation of the method used to ensure a good frequency response is given, and the transfer function of various parts of the circuit analysed. The experimental results are quoted as being similar in amplitude to a transfer of $1/(1 + 5 \times 10^{-4} j\omega)$, although the phase shift is considerably better than this.

A number of circuits used to protect the transistors against damaging conditions are discussed. Special attention is given to a separate transistor circuit which disconnects all the power supplies in the event of an irregularity in any one of them.

LIST OF SYMBOLS

A, B, C = Multiplier input channels.
A = Gain of amplifier.
C = Capacitance.
i = Instantaneous current.
I_{c0} = Collector leakage current for $V_B = +1$ volt.
k = Multiplicative factor.
L = Inductance.
p, q, r, s, t = Dimensions of attenuator waveform.
p = Instantaneous power.
P_{av} = Average power.
R = Resistance.
τ = Time-constant.
v = Alternating component of signal voltage, or instantaneous voltage.
V = Signal voltage.
V' = Constant voltage used for obtaining 4-quadrant operation.
V_0 = Output of error amplifier.
ω = Angular frequency.
Z = Impedance.

(1) INTRODUCTION

In the study of aircraft flight, it is necessary to simulate the aerodynamic conditions experienced by aircraft. When dealing with problems to which an accurate solution is required, it is necessary to include terms which involve the product of other terms. The evaluation of these products requires a multiplier with an accuracy within about 2% and capable of handling the signals encountered in this type of problem with negligible attenuation or phase shift. These signal waveforms are likely to contain components up to a frequency of 20 c/s, so that the formation of the squares of such waveforms will require a linear frequency response up to 40 c/s.

Mr. Gleghorn is with Smiths Aircraft Instruments, Ltd.

With the multipliers needed in simulating the aerodynamic conditions, together with those used in other parts of the problem, a typical analogue-computer installation may require up to twelve multipliers. It is therefore desirable that cost should be carefully considered in the design, and also that the setting-up procedure should involve the minimum of time, and preferably be done on a meter. It is also considered essential that the multiplier should be capable of 4-quadrant operation, i.e. operation in which both multiplicands may be of either sign, without the need for additional amplifiers.

A number of different types of multiplier are in current use, and these were considered with the above specification in mind. Those which depend on valve characteristics, i.e. parabolic I_a/V_g relationships, were considered too susceptible to changes in characteristics and would require considerable readjustment in the event of a valve change. Principles of multiplication which involve the use of an accurately defined waveform such as a triangular wave were not considered suitable. Since the final output of the multiplier unit is the difference of two voltages, owing to the method of obtaining 4-quadrant operation, it is necessary for the accuracy of the product-forming device to be considerably better than that required at the output, perhaps by a factor of 10. The generation of a waveform of this accuracy would require a fairly complicated circuit which would increase the cost of the unit. There would also be difficulties in the alternative method of feeding such a waveform from a central generator to the various multipliers without distortion.

More detailed consideration was given to the modulated square-wave type of multiplier* and especially to the use of transistors as modulators. A difficulty arises here because it is unusual in present practice to combine transistors and thermionic valves in the same circuit, owing to the risk of damage to the transistors; for in the event of failure of part of the thermionic-valve circuit, an excessive voltage might be applied to the transistor electrodes. Considerable attention was thus paid to the protection of the transistors, and the result of this work is discussed in Section 8.

(2) THE MODULATED SQUARE-WAVE TYPE OF MULTIPLIER

A block schematic of the modulated square-wave type of multiplier is shown in Fig. 1.

Two identical modulators have a transfer controlled by the mark/space ratio of a square wave, which is determined by the output of the error amplifier. Into this amplifier are fed one of the multiplicands and the output of modulator A. The results derived in Section 10.1 show this output to be

$$-V_A f(V_0) = k_1 V_C - \frac{k_2 V_0}{A} \quad . \quad . \quad . \quad . \quad (1)$$

* GREENWOOD, I. A., HOLDAM, J. V., and MACREA, D.: 'Electronic Instruments' (Radiation Laboratory Series No. 21, McGraw-Hill, 1948).
KORN, G. A., and KORN, T. M.: 'Electronic Analog Computors' (McGraw-Hill 2nd Edition, 1956).

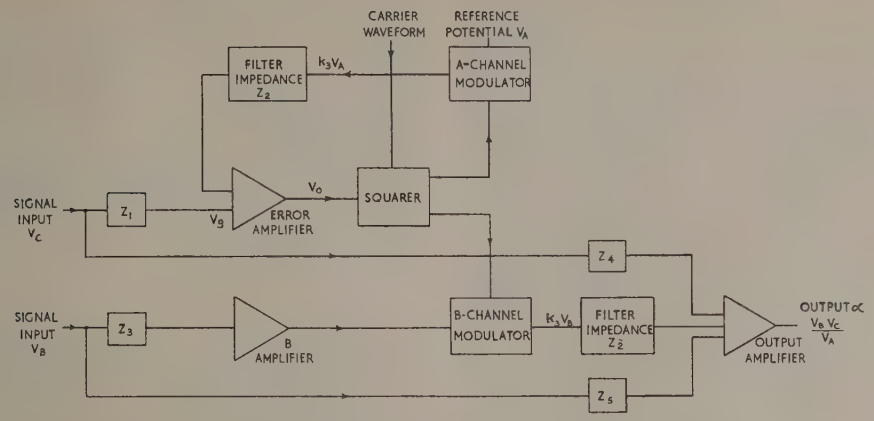


Fig. 1.—Schematic of modulated square-wave type of multiplier.

where k_1 and k_2 are the ratios of resistances and A is the gain of the error amplifier. The output of modulator B is

$$-V_B f(V_0) = \frac{V_B}{V_A}\left(k_1 V_C - \frac{k_2 V_0}{A}\right) \quad . \quad . \quad . \quad (2)$$

Since A is large, the term in V_0 can be neglected, and the output of the B-channel modulator can be regarded as proportional to the product $V_B V_C$. The term $f(V_0)$, which includes all the non-linearities associated with the relation between V_0 and the mark/space ratio, has been eliminated; this system therefore provides a high degree of linearity without undue complication.

A change of sign of either V_A or V_C is equivalent to changing the sign of the transfer of the modulator, which cannot be realized in practice. V_A and V_C must therefore always be of the same sign, additional circuits being included to permit 4-quadrant operation.

(3) THE MODULATOR

The modulator consists of a transistor in which the base is either driven negative, placing a very low impedance between the point X (Fig. 2) and earth, or positive, when the impedance is very high. In the ideal modulator, this impedance should be infinite; the degree of departure of the transistor from the ideal is determined by I_{c0}, which will continue to flow into the base and out of the collector when the transistor has been cut off by a positive voltage applied to the base. In any one transistor in a stable ambient temperature (a condition which can be reasonably assumed in a computer cabinet), I_{c0} can be considered as being affected only by the power dissipation in the transistor. As the voltage applied to the collector of the transistor modulator varies between the two extreme operating conditions, so the change in I_{c0} resulting from the change in power dissipation varies from $+10$ to $-3{\cdot}5\%$ of the I_{c0} flowing with zero B-channel input (see Section 10.2). This calculation is with $R_{23} = 6{\cdot}8$ kilohms. It should be noted that, after a change in the applied collector voltage, there will be a delay before the corresponding change in I_{c0} is complete; this delay will be of similar order to the period of the highest signal frequency.

With zero B-channel input, a typical I_{c0} (e.g. $2\,\mu$A) would cause a steady error at the output of the multiplier of 428 mV, having been amplified by 40 in the output amplifier. This can, however, be cancelled in the output stages. The $+10\%$ change referred to, resulting in an error of 43 mV at the multiplier output, is not linearly related to the signal, and can therefore be only approximately cancelled out. Because of the delay of the change in I_{c0} after the applied collector voltage is changed, this cancellation will become less effective as the signal frequency rises.

If R_{23} were doubled to reduce the power dissipation in the transistor, the typical I_{c0} would cause a steady error at the output of 856 mV. The 10% change in I_{c0} would now become 5%, but the resultant error would be the same, i.e. 43 mV, as

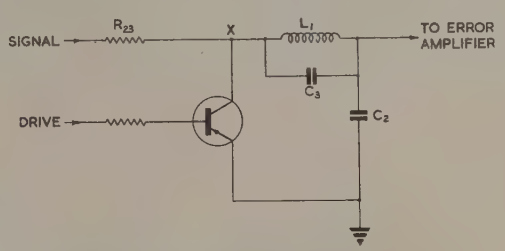


Fig. 2.—Transistor modulator.

before. R_{23} is thus chosen as providing the best compliance with other conditions.

In order that the modulation shall be accurate and the square wave at X undistorted, it is essential that the time-constant at X be kept low. The time-constant at a point is defined as being $C_E R_S R_L/(R_S + R_L)$, where C_E is the effective capacitance and R_L the resistive load between that point and earth and R_S is the output resistance of the signal source supplying C_E. In this case, R_L is assumed infinite when considering the wavefront of the square wave due to the presence of L_1, and the time-constant at X is thus $R_{23}(C_T + C_3)$, where C_T is the collector-emitter capacitance of the transistor. It is therefore necessary to keep R_{23} as small as is compatible with the available current output of the B-amplifier, thus fixing the value at $6{\cdot}8$ kilohms.

Because of the current amplification available in the transistor, only a small current is required to drive it into the fully conducting state, and it can therefore be driven directly from the squarer via a passive network. The most suitable type of transistor was found to be one recently developed for switching applications, the type OC.76.

(4) THE MODULATOR DRIVE

Fig. 1 shows the two modulators whose transfers are controlled by the mark/space ratio of a square wave. This square wave is obtained from a 2-stage over-driven amplifier with a sine-wave input, in which the bias, and hence the point at which the output

is limited owing to the flow of grid current, is set by the error amplifier. The proportion of the sine wave thus limited will determine the output mark/space ratio. Since the amplitude of the sine wave is much greater than the grid base, the squarer will be substantially independent of valve characteristics. The relationship of the mark/space ratio to the output of the error amplifier will be a sine function, but its departure from linearity will be divided by the gain of the error amplifier when the negative feedback loop is closed.

The overall frequency response of the modulator and its associated filter will be primarily governed by the frequency of the modulator drive or carrier. This should be high compared with the maximum signal frequency to ensure accurate modulation, and also to allow the tuned filter to attenuate the carrier adequately without shifting the phase of the maximum signal frequency. If, however, the drive frequency is too high, there will be difficulties in producing an accurate square waveform (which will include components up to 10 times the carrier frequency) with a transistor, whilst keeping the circuit simple and relatively unaffected by stray capacitances. Consideration of these factors led to the adoption of a 10 kc/s modulator drive frequency.

(5) 4-QUADRANT OPERATION

The multiplier, so far, has been capable of operation with inputs each of only one sign. V_A and V_B must be negative to give a negative transistor collector voltage, and therefore V_C must be positive. For simplicity, V_A was fixed, restricting the multiplier to product-forming and squaring, and its scope was then extended to 4-quadrant operation so that both inputs could be of either sign.

This was achieved by providing an input amplifier in the B- and C-channels which added a constant voltage, V_B' or V_C', to the signal and attenuated the sum, so that, for the normal range of input signals, ± 50 volts, the output of the input amplifier was of the correct sign and restricted to the limits of operation of the channel concerned. A separate B-amplifier is used to perform this function in the B-channel, but the functions of the C-channel input amplifier are performed by the error amplifier. The complete circuit diagram of the multiplier is shown in Fig. 3.

(6) ACCURACY WITH DIRECT-VOLTAGE INPUTS

The chief factors affecting the d.c. accuracy of the multiplier are:

(*a*) The linearity with reference to V_C of the mark/space ratio of the square wave applied to the modulators and hence of the modulated output.

(*b*) The similarity of the two modulators, which in the calculations in connection with (*a*) are assumed identical.

(*c*) The drift stability of all amplifiers, the accuracy of their transfers, and the accuracy of the various cancellation voltages fed into the output amplifier.

The effect of distortion of the waveform at the transistor collector due to capacitance to earth at this point and to a distorted drive waveform will also be examined.

The exact expressions for (*a*) are derived in Section 10.1. The error term in the linearity of the relationship is a function of V_0/A, where V_0 is the error amplifier output and A its gain. The sinusoidal nature of the carrier input to the squarer is contained in the V_0 term, and any non-linearity of the amplifier gain with input in the A term. Substituting experimental values as in Section 10.1 into this expression suggests that the error term is less than 0·1% of V_C.

The similarity of the two modulators can be examined in relation to three quantities, namely the open-circuit resistance of the cut-off transistor, the short-circuit resistance of the fully conducting transistor and the collector-to-emitter capacitance, C_T. The first of these quantities is so large and the second so small compared with the external impedance at the transistor collector that even quite large variations in them can be tolerated without affecting the transfer characteristic of the modulator. The third factor causes distortion of the waveform at the transistor collector. Owing to the hole-storage effect, this capacitance is considerably greater at low collector voltages, and thus has the greatest effect when the transistor has just been cut off and the collector voltage is starting to rise; the time-constant at the collector is $R_{23}(C_T + C_3)$, which at this point in the cycle might be 5×10^{-6} sec. However, of this, 70–80% is due to the effect of C_3, which therefore considerably reduces the effect of the difference in C_T between the transistors forming the two modulators. The effect of the lower C_T applying at higher

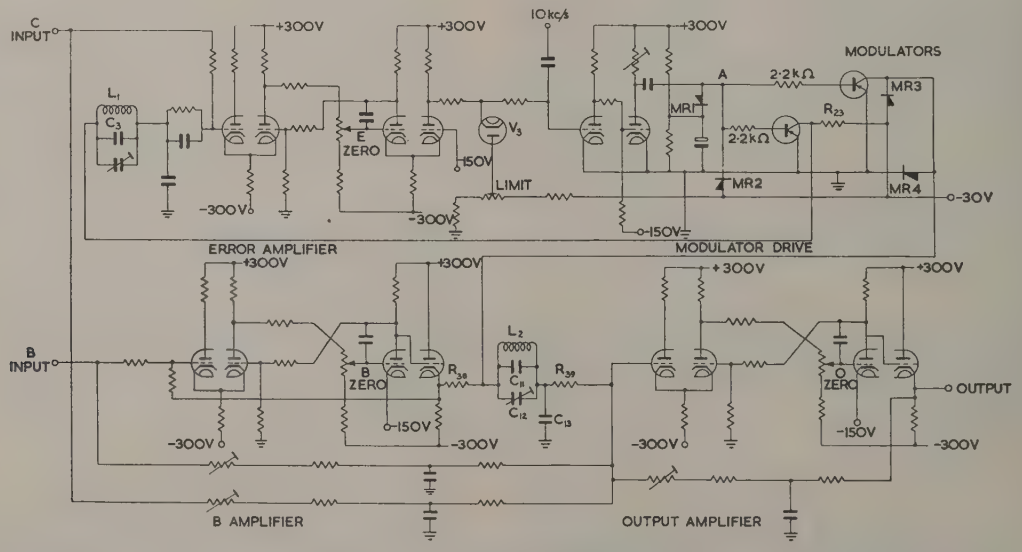


Fig. 3.—Circuit diagram of multiplier unit.

collector voltages will be correspondingly less. The matching of the two transistors would therefore appear to be of secondary importance, and this is confirmed in practice.

It is interesting at this point to consider the effect of distortion at the transistor collector (which is primarily due to distortion in the modulator drive) on the accuracy of the multiplication. The form of the modulated signal is shown in Fig. 4, which will be considered as being approximately bounded by straight lines.

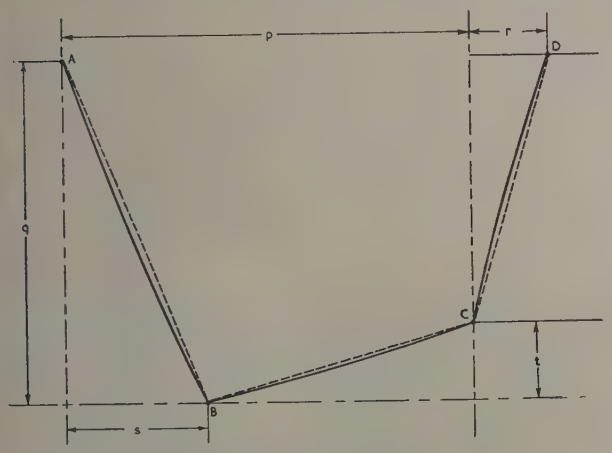


Fig. 4.—Modulator waveform.
—— Actual.
---- Approximated.
—·— Ideal.

In the case of the A-channel modulator, only the dimension p will vary, whereas in the B-channel, p and q will both vary. Since the form of the modulator output is mainly due to the form of the drive from the squarer, the dimensions r and s are independent of p and q, whereas t is proportional to q and approximately independent of p. If $t = kq$, the area can be expressed as

$$q\left[p\left(1-\frac{k}{2}\right)+\frac{1}{2}(r-s)(1-k)\right] \quad . \quad . \quad (3)$$

Since p is the only variable in the A-channel modulator, the area is proportional to pq plus a constant term. Since this constant term can be compensated for in the setting of V_C', the departure of the waveform from the ideal does not affect the correct operation.

In the B-channel modulator, the proportionality is maintained, plus an error term dependent only on q, i.e. on V_B; this can similarly be compensated for in the setting of the $V_B V_C'$ term.

If the sides of the figure lie between the same four points A, B, C, D, but are represented by the actual curve instead of the approximated curve, the above calculations are still valid, since only a number of constants of proportionality will be introduced. Hence, provided that the waveform can be considered as being reasonably similar to the one shown in Fig. 4, its shape will not affect the accuracy of the multiplication.

The three d.c. amplifiers used in the multiplier are of the non-drift-corrected type, having an open-loop gain of about 2000, since it had been decided to tolerate the effects of drift in order to reduce the cost of the unit. The short-term drift at the output of the multiplier with the inputs earthed is ±20 mV, and the long-term, ±200 mV, measured over periods of 1 min and 1½ h respectively. Most of this drift is due to the high gain of the output amplifier, which for drift calculations is about 55, i.e. feedback resistance divided by the parallel combination of all the input resistances.

After setting the multiplier to give zero output with earthed inputs and correctly adjusting all the variable resistors setting the balancing voltages, a series of measurements of output voltage was made for various input voltages evenly spread over the entire range of operation on the two channels. The overall error was found to be ±2%, although many of the points were within ±1%. The drift in the zeros of the amplifiers was checked and allowed for before taking each reading to avoid confusing the measurements.

In order to obtain a maximum output accuracy of within 1% at the output of the multiplier, it is necessary for the modulator to be accurate to better than 0·1%. This is, however, the order of errors caused by the presence of the V_0/A term, and it is not possible to justify to this accuracy some of the approximations made in considering the effect of waveform distortion at the transistor collector. Improvements in the modulator drive circuit and drift-compensation of the error and output amplifiers would probably increase the accuracy of the multiplier, but this would have to be offset against the increased size and cost of the unit.

(7) FREQUENCY RESPONSE

The filter following the modulator is shown in Fig. 3. It comprises L_2, C_{11}, C_{12} and C_{13}; the anti-resonant frequency determined by L_2 C_{11} is at the carrier frequency of 10 kc/s, and the resonant frequency determined by L_2 C_{13} is at 1 420 c/s. The effect of C_{11} is neglected in the signal-frequency analysis.

If V_C is fixed and an alternating voltage V_B applied to the filter by the B-amplifier, the voltage at P (the junction of C_{13} and R_{39} in Fig. 3) will be

$$V_P = \frac{0\cdot787 V_B k_3}{1 - 9\cdot85 \times 10^{-9}\omega^2 + j\omega 1\cdot49 \times 10^{-4}} \quad . \quad (4)$$

(see Section 10.3). It is arranged that the B by-pass channel has a similar transfer to ensure correct cancellation. The overall lag in the modulated and by-pass channels is corrected by a phase advance network associated with the output amplifier.

The analysis of the transfer of the C-channel is complicated by the sine function introduced by the squaring circuit, which causes the transfer to be a function of V_0. The result is, however, similar.

The various correcting networks described enable the amplitude of the output of the multiplier to be maintained within 1·5% of its d.c. value at 50 c/s, and within 4% at 100 c/s, equivalent in amplitude to a transfer of $1/(1 + 5 \times 10^{-4} j\omega)$; however, the phase rotation at these frequencies is considerably better than this transfer suggests, being within 1·7 and 2·8° respectively.

There is an output from the multiplier when one input is earthed and a signal applied to the other; at zero frequency it is of the order of ½–1% of the modulus of the applied signal. Its magnitude does not change up to about 30 c/s; at 50 c/s it is 1% and at 100 c/s 1·5%.

Under all conditions of input, a residual 10 kc/s carrier of about 5 mV r.m.s. amplitude could be observed at the output of the multiplier.

(8) PROTECTIVE CIRCUITS

The normal signal range of V_C is ∓50 volts, and the corresponding range of V_0, 6·9 ± 9·6 volts. If these limits of V_0 are exceeded, the square-wave drive to the modulators will be inhibited, as the sine-wave carrier will cease to intersect the working range of the grid of the first squaring amplifier. If V_0 rises positively, the mark/space ratio will increase, thus

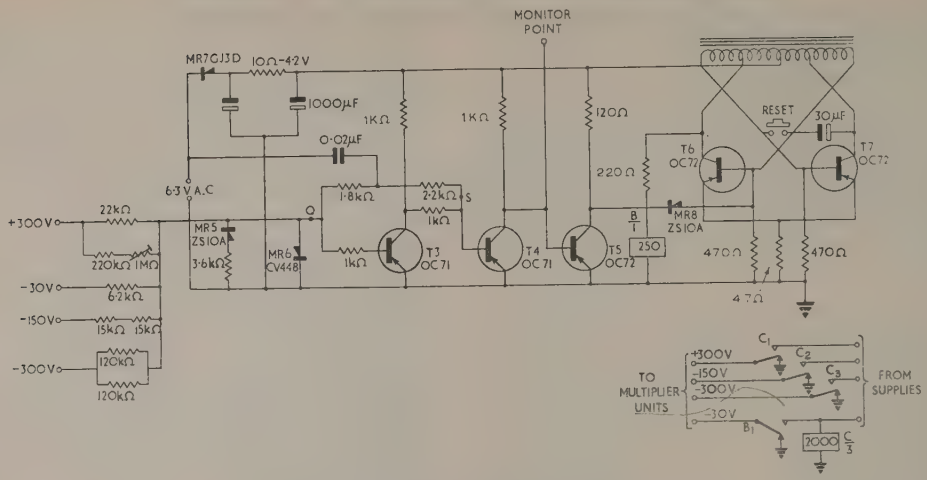


Fig. 5.—Transistor detection circuit.

increasing the effect of V_A on the error amplifier. As the limit of V_0 is exceeded, however, the amplitude of the drive waveform will be reduced; since the drive is capacitively coupled to the modulator, this will prevent it being fully opened. Under these conditions, the effect of V_A on the error amplifier will be an inverse function of V_0, which will constitute positive feedback. This will prevent a restoration of the drive waveform when V_0 regains its normal working range, and a diode V_3 was therefore included at the output of the error amplifier to prevent this limit of V_0 being exceeded.

One of the chief difficulties of using transistors associated with thermionic circuits is in protecting them satisfactorily in the event of any failure of such associated circuits or their power supplies. Protection in this multiplier is based on the provision of a stabilized −30-volt supply, which can be rapidly earthed, and a detection circuit to initiate this rapid earthing and to switch off all other power supplies.

It is desirable that the following limitations should be applied to the transistor electrodes:

(*a*) The applied collector voltage should be between zero and −30 volts.

(*b*) The base voltage should not exceed +1·5 volts.

(*c*) The base current and collector dissipation should not rise excessively.

Referring to Fig. 3, it will be seen that the collector of the A-channel modulator will comply with condition (*a*) to the extent that the −30-volt supply complies; this is discussed later. Condition (*b*) is applied to both transistors by the same diode MR1; any positive rise of voltage during the warming-up period at the common drive point A above the fixed bias determined by the resistor chain is prevented by this diode. A negative-going pulse at this point, such as might occur when the unit is switched on, is limited to the potential of the −30-volt supply by diode MR2. This, in conjunction with the base-drive resistance of 2·2 kilohms, limits the base current to −13·7 mA, thus easily satisfying condition (*c*).

The diode MR3 protects the collector of the B-channel modulator against voltages exceeding −30 volts, and diode MR4 against positive voltages, thus complying with condition (*a*). Conditions (*b*) and (*c*) are dealt with by circuits which are common to both channels.

On the separate chassis which contains the −30-volt stabilized supply is the transistor circuit for the detection of voltage changes in any of the power supplies (Fig. 5). A resistor chain connecting all the power supplies has its central point Q about ½ volt negative with respect to earth; this voltage can be adjusted by means of a variable resistor. The negative voltage at Q, V_Q, is applied to the first stage of the detection circuit. This is a sign-reversing amplifier T_3 with a series output 1-kilohm resistor, the whole stage being in parallel with a 4-kilohm resistor. V_Q is adjusted so as to increase the collector current of T_3 until saturation conditions are just reached. If V_Q then becomes less negative due to a change in the voltage of one of the power supplies, the collector of T_3 will become more negative as the transistor is cut off. Since changes at the collector of T_3 will have four times the effect of changes at Q on the point S, due to the ratio of the two series resistors, this will cause the base of T_4 to become more negative, and hence its collector less negative.

If V_Q, however, becomes more negative, no further increase in the collector current of T_3 can occur, and hence its collector voltage will not change. Thus the voltage of S will follow V_Q at one-fifth of the rate, and the collector of T_4 will become less negative as in the previous case. The initial setting of V_Q is made by adjusting for a maximum negative voltage at the monitor point on the collector of T_4; any change in V_Q, which will then cause this point to become less negative, is fed via the heavy-current sign-reversing stage T_5 and a diode MR8 to the bistable multivibrator, T_6 and T_7.

This multivibrator has its stable conditions determined by the resistance of the various sections of the transformer winding, and the change-over between stable states is assisted by the positive feedback provided by the transformer coupling. A sensitive high-speed relay, B, is connected between the collector of one of the transistors and earth, and is energized in the 'ready' condition of the detection circuit. This provides a fail-to-safety feature in the event of failure of the relay or of the 6·3-volt a.c. supply. When the multivibrator is triggered by a change in one of the main power supplies affecting V_Q, the relay B will be de-energized, disconnecting the −30-volt line from its supply and earthing it. Whilst the contacts of relay B are changing, the −30-volt line will remain referred to earth through the 2-kilohm coil of relay C. This relay disconnects all power supplies to the multipliers if the −30-volt supply fails. It will be seen that if one of the supplies to the multiplier fails, attempting to impose excessive conditions on the transistors, protection will be afforded by earthing the −30-volt line until relay C has had time to operate. The speed of operation of the whole chain is found sufficient to protect the transistors; in the event of a

complete failure of one of the supplies, the −30-volt line will be disconnected in about 1¼ millisec and earthed in about 6 millisec.

It is, however, possible that, instead of failing completely, one of the main power supplies may change slightly from its nominal setting. From these supplies are derived a number of balancing voltages, a change in one of which could cause the maximum transistor ratings to be exceeded. Hence the detection circuit should be capable of responding to very slow changes in V_Q. Since the multivibrator will respond only to a steeply rising wavefront, a pulse train is injected into the channel in the first stage of the detection circuit. When T_3 is in the linear region, this is provided by the ripple from the transistor power supply, and when in the saturated current region, by an injected voltage via a 0·02 μF capacitor from the 6·3-volt 50 c/s supply. Amplification and limiting in the subsequent stages shape these injected components sufficiently to enable them to trigger the multivibrator. Whilst the detection circuit is in the ready condition, these triggering pulses lie below the operational threshold of the multivibrator, but a change in V_Q will cause them to project into the operating region, and initiate the change between stable states. The device will detect changes of 1 volt in any of the power supplies.

To prevent T_3 from being damaged by large excursions of V_Q in the event of complete failure of a power supply, the diode MR5 is included. Although biased in the foward direction by the negative voltage, V_Q, this is below the knee voltage for the type ZS10A diode, and the diode resistance is therefore high. If V_Q becomes more negative, the diode will conduct; the series 3·6-kilohm resistance will allow V_Q to make an excursion which is ample to ensure triggering but which will not exceed a few volts if a power supply fails. A diode, MR6, in the reverse direction prevents V_Q from becoming positive with respect to earth.

The combination of these protective devices permits the satisfactory interworking of transistor and thermionic circuits; there have been no instances of transistor failure with the protective devices in use, although they did occur before their introduction.

(9) ACKNOWLEDGMENTS

The author is indebted to Mr. K. Fearnside, the Research Manager of Smiths Aircraft Instruments, Ltd., for permission to publish this work, and also wishes to acknowledge many helpful discussions with Mr. J. Evans of the Experimental Research Unit.

(10) APPENDICES

(10.1) Transfer of the Error Amplifier with Direct-Voltage Input

Referring to Fig. 1,

$$\frac{V_C - V_g}{Z_1} + \frac{V_A f(V_0) - V_g}{Z_2} = 0 \quad . \quad . \quad . \quad (5)$$

$$V_g = \frac{V_0}{A} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Therefore $$-V_A f(V_0) = V_C \frac{Z_2}{Z_1} - V_0 \frac{Z_1 + Z_2}{AZ_1} \quad . \quad . \quad . \quad (7)$$

Let $k_1 = Z_2/Z_1$ and $k_2 = (Z_1 + Z_2)/Z_1$.

Then $$-V_A f(V_0) = k_1 V_C - k_2 V_0/A \quad . \quad . \quad . \quad . \quad (8)$$

Since both modulators have the same transfer, the output of the B-channel modulator will be

$$-V_B f(V_0) = \frac{V_B}{V_A}\left(k_1 V_C - \frac{k_2 V_0}{A}\right) \quad . \quad . \quad . \quad (9)$$

Substituting values into the term k_2/A gives an error term of about 0·05% of V_0. The relative orders of magnitude of V_0 and V_C suggest that this error will be about 0·1% of V_C.

(10.2) The Relationship of I_{c0} and Applied Collector Voltage to the Transistor Modulator

It is assumed that power is dissipated in the transistor only during the change between states and not when the transistor is fully cut off or fully conducting. If the voltage at the transistor collector, whose instantaneous value is v, rises linearly to a final value V as the transistor is cut off, the instantaneous power dissipation in the collector can be expressed as

$$p = vi = \frac{Vv - v^2}{R_{23}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where i, the current flowing out of the transistor, is equal to $(V - v)/R_{23}$. The average power over the rise is then given by

$$P_{avr} = \frac{1}{V}\int_0^V \left(\frac{Vv - v^2}{R_{23}}\right) dv = \frac{V^2}{6R_{23}} \quad . \quad . \quad (11)$$

The transistor is occupied for one-sixth of the cycle in changing between fully conducting and fully non-conducting and back again. The average power over the whole of the cycle is thus

$$P_{avc} = \frac{V^2}{36R_{23}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

When $V = 15$ volts, $P_{avc} = 0·92$ mW (i.e. the condition for zero B-channel input) and when $V = 30$ volts, $P_{avc} = 3·68$ mW. Assuming that the base power is independent of V, the change in the transistor power dissipation between the two extremes of its operating conditions is from +2·76 to −0·92 mW. At a temperature coefficient of 0·4° C/mW, this will result in changes in the transistor temperature of +1·10 and −0·368° C, and resulting changes in I_{c0} of +10 and −3·5%.

(10.3) Transfer of the Modulator Filter

For the purpose of this analysis, the transistor modulator is assumed to reduce the input v_B to $v_B f(V_0)$ and to have no other effect on the signal, since signal-frequency components occur throughout the network and are not strictly modulated by the carrier waveform. The effects at signal frequency of the collector-to-emitter capacitance of the transistor modulator and of the capacitance C_{11} shown in Fig. 3 are neglected.

The term $f(V_0)$ will be written as a multiplicative factor k_3, which is a constant when V_c is constant. The voltage at P, the junction of C_{13} and R_{39} in Fig. 3, is given by

$$V_P = \frac{k_3 v_B R_{39}/(1 + j\omega C_{13} R_{39})}{R_{38} + j\omega L_2 + R_{39}/(1 + j\omega C_{13} R_{39})} \quad . \quad . \quad (13)$$

$$= \frac{k_3 v_B}{1 + k_4}\left[\frac{1}{1 - \omega^2 k_4 \tau_1 \tau_2/(1 + k_4)}\right] \times \left[\frac{1}{1 + j\omega \dfrac{(\tau_1 + \tau_2)/(1 + k_4)}{1 - \omega^2 k_4 \tau_1 \tau_2/(1 + k_4)}}\right] \quad . \quad . \quad (14)$$

where $\tau_1 = C_{13} R_{39} = 6·25 \times 10^{-4}$ sec, $\tau_2 = L_2/R_{38} = 7·36 \times 10^{-5}$ sec and $k_4 = R_{38}/R_{39} = 0·272$.

Therefore

$$V_P = \frac{0·787 k_3 v_B}{1 - 9·85 \times 10^{-9}\omega^2 + j\omega 1·49 \times 10^{-4}} \quad . \quad (15)$$

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL SECTION, 3RD NOVEMBER, 1959

Mr. G. Herring: A great deal has been said in the past about high-accuracy multipliers, and so it is pleasing to see that the applications where only modest accuracy is required are also being recognized.

This multiplier must be considered in the light of other techniques available. One which has not been mentioned is the servo multiplier, which can provide a satisfactory performance, together with the facility for multiple multiplication, if one of the inputs has a bandwidth of less than about 10 c/s. With present-day techniques the size may be even less than for the device described. Another possibility is to use the quarter-squares principle. Essentially this relies on producing a voltage proportional to the square of another voltage by using a computing amplifier with a non-linear input resistor. One way of obtaining a non-linear resistance is to use a group of diodes biased by resistors operating from a reference potential to produce the required law. However, this can be bulky and expensive even with semiconductor diodes, and an alternative where only moderate accuracy is required is to use a resistor with an inherently non-linear variation of current with applied voltage. By using trimming and padding resistors a parabolic characteristic can be produced, so that this type could be a contender, although four computing amplifiers are needed for each full multiplication.

To come to this particular multiplier, the advantages of the feedback method of generating the square wave are perhaps not stressed sufficiently in the paper. One feature which makes it a very attractive scheme is that the error amplifier has to handle only low-frequency signals. Other mark/space ratio multipliers require that the amplifiers shall handle triangular or square waves.

On the question of transistor modulators, we have recently developed some switch circuits at the Royal Aircraft Establishment which can handle positive and negative input voltages. Although we were only interested in inputs in the range ± 5 volts at the time, there is nothing inherent in the design which would prevent their being used up to twice this voltage. The speed and switching accuracy achieved may be within the requirements for this multiplier and a switch capable of handling both positive and negative signals would eliminate the need for the biasing voltages in the present multiplier.

Mr. H. Faulkner: The author has produced a multiplier in the medium accuracy range of 1–2% which is quite necessary in the computing field for aerodynamics, where several cross-product quantities are used.

On the question of accuracy, 1% for most of the range is fairly reasonable. This has been obtained mainly because of the feedback on the mark/space squaring device. Other types of multiplier go to great lengths to obtain a linear mark/space ratio with voltage input, and do not adopt the simpler course of putting feedback round the loop.

With reference to temperature, are the two transistor modulators matched? It is possible for the ambient temperature in computer racks to change between 5 and 10° C between 8 a.m. and 10 a.m., especially in laboratories, etc., where space heating is switched off over night, and this may lead to a great deal of drift in the warming-up period. Is it necessary during the setting-up of the multipliers to match the transistors used in the modulator for I_{co}/temperature characteristics? A 10° C change may cause I_{co} to change by a factor of 3.

Can the author give any information on the effect of drift of the 30-volt reference power supply? Any drift could cause errors, and it may be that these errors could be worse with one input at zero and the other input at maximum.

Most 4-quadrant multipliers give the facilities of division and square-root extraction. Can the author comment on whether it is possible to modify the multiplier for this purpose?

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. P. Gleghorn (*in reply*): In reply to Mr. Herring, I agree that the quarter-squares principle may be used with an input resistor in which the current does not vary linearly with the applied voltage and that this could lead to quite a small multiplier. However, this type of resistor frequently depends for its characteristic on a change of resistance with temperature, and I therefore feel that the long-term stability of such a device should be carefully checked, especially if the setting up of the trimming and padding resistors were to involve a lengthy process of alternate adjustment and transfer plotting. Also, the effects of the thermal time-constant should be carefully considered.

Mr. Herring's remarks about the possibility of transistor switches capable of accepting either sign of input are most interesting. If these proved suitable, the B-channel amplifier could be dispensed with, and the opportunity taken to redesign the squaring circuit to use transistors instead of thermionic valves, which would enable the protection arrangements to be simplified.

The matching of transistors for I_{co}/temperature characteristics mentioned by Mr. Faulkner could not be done simply, as, in the one case, variations in I_{co} would be equivalent to changes in V_A, the divisor, and, in the other, to changes in V_B, the dividend. The multiplier was originally intended to be used in computers where the power was on continuously, thus reducing the effect of changes in the external surroundings.

Drift in the -30-volt reference supply will affect directly the input, V_A, and hence the output. Section 5 refers to the addition of the inputs V_B' and V_C', so that the output of the B-channel modulator is proportional to $(V_B + V_B')(V_C + V_C')/V_A$. The steady error $V_B'V_C'/V_A$ is cancelled by a voltage fed into the output amplifier, and the errors proportional to V_B and V_C by signals fed via the by-pass channels, Z_5 and Z_4 (Fig. 1) respectively. If, therefore, the voltage V_A drifts, not only will the product term V_BV_C/V_A be in error, but there will be incomplete cancellation of the unwanted terms which may give rise to considerable errors. This is the reason for using a highly stabilized source for this supply.

Division would be effected by varying V_A as the divisor, but this would lead to the cancellation errors mentioned above, and would therefore be impracticable. It may be possible to connect the multiplier in the feedback loop of a computing amplifier in order to perform a division.

621.375.9 : 621.372.2

The Institution of Electrical Engineers
Paper No. 2991 E
Aug. 1959
©

THEORY OF THE TRAVELLING-WAVE PARAMETRIC AMPLIFIER

By Professor A. L. CULLEN, Ph.D., Member.

(*The paper was received 7th January*, 1959. *It was published in August*, 1959, *and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION *9th November*, 1959.)

SUMMARY

The theory of a parametric travelling-wave amplifier is developed, including the pumping wave explicitly.

Particular reference is made to two characteristics of such amplifiers which theories assuming a constant pumping-wave amplitude fail to predict.

The first arises from ohmic attenuation in the line, and limits the useful length of line to a length of the order of $(\alpha_1\alpha_3)^{-1/2}$, where α_1 and α_3 are the signal- and pumping-wave attenuation coefficients. Further increase in length leads to a decrease in gain.

The second characteristic is a saturation effect at high signal levels. This is found to be associated with a spatially periodic power transfer between signal and idling waves on the one hand and pumping wave on the other. The maximum power output at the signal frequency is obtainable from the input pump power and the input signal power through the Manley–Rowe relations.

LIST OF PRINCIPAL SYMBOLS

C = Capacitance per unit length, a function of voltage.
C_0 = Small-signal value of C.
u = Small-signal phase velocity.
G = Conductance per unit length
i = Instantaneous current.
l = Length of line.
L_0 = Inductance per unit length.
n = An integer.
P_1, P_2, P_3 = Power in signal, idling or pumping waves.
R = Series resistance of crystal diode.
t = Time.
v = Instantaneous voltage.
x = Distance along line.
Z_0 = Characteristic impedance of line.
$\alpha_1, \alpha_2, \alpha_3$ = Attenuation coefficient of signal, idling and pumping waves due to ohmic losses.
α, α_0 = Growth constants of amplified waves.
$\beta_1, \beta_2, \beta_3$ = Phase coefficients of signal, idling and pumping waves.
η = Constant of proportionality between capacitance change and voltage.
$\phi_1, \phi_2, \phi_3, \psi$ = Phase angles.
$\omega_1, \omega_2, \omega_3$ = Angular frequencies of signal, idling and pumping waves.

(1) INTRODUCTION

The theory of the travelling-wave parametric amplifier has been discussed previously by several authors.[1–3] The present treatment differs from previously published accounts in several respects.

In the first place, the starting-point of previous theories is a travelling wave of distributed inductance or capacitance, whilst the present paper assumes initially a non-linear distributed capacitance excited by a travelling wave of pumping voltage.

Prof. Cullen is Professor of Electrical Engineering, University of Sheffield.

Secondly, the effect of attenuation due to series-resistance losses associated with the capacitance is studied.

Thirdly, consideration is given to the effect of saturation at large signal levels.

A new mechanism of periodic transfer of power between the various waves is found.

(2) BASIC THEORY

The practical form of travelling-wave parametric amplifier which we have in mind is a series of low-pass filter sections with non-linear shunt capacitance, provided, for example, by suitably biased silicon junction diodes. In an alternative practical arrangement, the diodes can be connected at equal intervals as shunt loads along a transmission line. A uniform coaxial line or a shielded helix are two possible forms which the transmission line could take.

The idealized form of travelling-wave parametric amplifier considered in the paper is a uniform transmission line having a constant inductance per unit length and a variable capacitance per unit length.

The capacitance per unit length is assumed to depend at any point along the line on the voltage at that point, according to the simple relationship

$$C = C_0(1 + \eta v) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It is further assumed that only three frequencies need be considered in analysing the performance of the line. These are the signal frequency, ω_1, the idling frequency, ω_2, and the pump frequency, ω_3. This assumption is necessary if the analysis is to be reasonably tractable, and in certain circumstances can be justified theoretically if the pumping voltage is much larger than the signal voltage and the idling voltage, at least when the capacitance/voltage law is of the simple form described by eqn. (1). The frequencies ω_1, ω_2 and ω_3 are related as usual by

$$\omega_1 + \omega_2 = \omega_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

We shall also neglect ohmic losses in the transmission line in the first instance, though attention will be given to this point later.

The transmission line equations, when C is variable, can be written

$$-\frac{\partial i}{\partial x} = \frac{\partial}{\partial t}(Cv) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$-\frac{\partial v}{\partial x} = L_0\frac{di}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Eliminating i from eqns. (3) and (4) yields the non-linear wave equation

$$\frac{\partial^2 v}{\partial x^2} = \frac{1}{u^2}\frac{\partial^2}{\partial t^2}(v + \eta v^2) \quad . \quad . \quad . \quad . \quad (5)$$

where

$$u = \frac{1}{\sqrt{(L_0C_0)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

We next assume that v can be written in the form

$$v = V_1 \cos(\omega_1 t - \beta_1 x + \phi_1) + V_2 \cos(\omega_2 t - \beta_2 x + \phi_2) + V_3 \cos(\omega_3 t - \beta_3 x + \phi_3) = v_1 + v_2 + v_3 \text{ say} \quad (7)$$

in which V_1, V_2 and V_3 depend only on x. The phase angles ϕ_1, ϕ_2 and ϕ_3 are arbitrary constants at this stage. The phase coefficients β_1, β_2, β_3 are assumed to have the small-signal values appropriate to the frequencies concerned. Thus $\beta = \omega/u$ in general, and in particular the β's satisfy

$$\beta_1 + \beta_2 = \beta_3 \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Eqn. (7) can be written in an abbreviated notation thus:

$$v = \sum_{n=1}^{3} V_n \cos \theta_n \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where

$$\theta_n = \omega_n t - \beta_n x + \phi_n \quad . \quad . \quad . \quad . \quad (9a)$$

Now consider the term v^2 which occurs in eqn. (5). We have

$$v^2 = v_1^2 + v_2^2 + v_3^2 + 2(v_1 v_2 + v_2 v_3 + v_3 v_1) \quad . \quad (10)$$

The six terms in this equation contribute respectively the following frequencies:

0 and $2\omega_1$, 0 and $2\omega_2$, 0 and $2\omega_3$, $\omega_1 + \omega_2$ and $\omega_2 - \omega_1$, $\omega_2 + \omega_3$ and $\omega_3 - \omega_1$, $\omega_3 + \omega_1$ and $\omega_3 - \omega_1$.

The zero-frequency terms we may ignore. Of the rest, $\omega_1 + \omega_2 = \omega_3$, $\omega_3 - \omega_2 = \omega_1$, and $\omega_3 - \omega_1 = \omega_2$, so that these terms can clearly be accounted for in eqn. (7). In assessing the importance of the remaining terms, we have to take into account the frequency and the amplitude of each term. Thus, whilst terms like v_1^2 and v_2^2 may justifiably be neglected in a first-order small-signal theory, it is preferable to retain the term of frequency $\omega_1 + \omega_2$ arising from the product $v_1 v_2$. The significance of this small term arises from the fact that its frequency is that of the pumping wave, by eqn. (2). If it is ignored, the resulting equations fail to conserve power. The term of frequency $\omega_2 - \omega_1$ arising from $v_1 v_2$ is, however, neglected. The remaining terms all involve v_3, and cannot be neglected on grounds of small amplitude. However they are all of frequencies higher than ω_3, namely $2\omega_3$, $\omega_3 + \omega_1$ and $\omega_3 + \omega_2$, and if it is assumed that a low-pass filter of cut-off frequency ω_3 is introduced appropriately at the termination, such frequencies will not appear in the output. Of course, they may have undesirable effects on the frequencies that do appear, particularly if they lead to instability and oscillation. Such effects are not considered here. They may well be important in practice, of course, but the increased complication of analysis needed to allow for them seems hardly justifiable at present, particularly in view of the highly over-simplified capacitance variation law which has been assumed.

When these various terms are neglected, we retain as the significant part of v^2, written sig (v^2), the following:

$$\text{sig}(v^2) = V_1 V_2 \cos(\theta_1 + \theta_2) + V_2 V_3 \cos(\theta_3 - \theta_2) + V_3 V_1 \cos(\theta_3 - \theta_1)$$

This can be written

$$\text{sig}(v^2) = V_1 V_2 \cos(\theta_3 + \psi) + V_2 V_3 \cos(\theta_1 - \psi) + V_3 V_1 \cos(\theta_2 - \psi) \quad (11)$$

in which the relationship $\beta_1 + \beta_2 = \beta_3$ has again been employed, and in which ψ is given by

$$\psi = \phi_1 + \phi_2 - \phi_3 \quad . \quad . \quad . \quad . \quad (12)$$

We next substitute eqns. (7) and (11) in eqn. (5), remembering that V_1, V_2 and V_3 depend on x.

When this is done, we can separate out terms of frequencies ω_1, ω_2 and ω_3, and so we get three independent equations. Each of these three equations is further separable, since components in phase quadrature must be separated. Hence six equations are obtained. Using the relationship $\beta_n = \omega_n/u$, these equations can be written as follows:

$$\frac{d^2 V_1}{dx^2} = -\eta \beta_1^2 V_2 V_3 \cos\psi \quad . \quad . \quad . \quad (13)$$

$$\frac{dV_1}{dx} = -\frac{\eta}{2} \beta_1 V_2 V_3 \sin\psi \quad . \quad . \quad . \quad . \quad (14)$$

$$\frac{d^2 V_2}{dx^2} = -\eta \beta_2^2 V_3 V_1 \cos\psi \quad . \quad . \quad . \quad (15)$$

$$\frac{dV_2}{dx} = -\frac{\eta}{2} \beta_2 V_3 V_1 \sin\psi \quad . \quad . \quad . \quad . \quad (16)$$

$$\frac{d^2 V_3}{dx^2} = -\eta \beta_3^2 V_1 V_2 \cos\psi \quad . \quad . \quad . \quad (17)$$

$$\frac{dV_3}{dx} = +\frac{\eta}{2} \beta_3 V_1 V_2 \sin\psi \quad . \quad . \quad . \quad . \quad (18)$$

A general solution of these equations will not be attempted, indeed it may not exist, for the equations are themselves only approximate, and we have not shown them to be mutually consistent. An approximate solution is, however, given in Section 3.

(3) LOSS-FREE SMALL-SIGNAL THEORY

To obtain an approximate solution we make the following two additional assumptions:

(*a*) Very small signal conditions; assume V_3 constant since very little power is extracted from the pumping wave. Note that this is consistent with eqn. (18), since dV_3/dx is of the second order of small quantities if V_1 and V_2 are both very small. We accordingly write $V_3 = V_{30}$, a constant.

(*b*) Small gain per unit length, so that $\frac{d^2V}{dx^2} \ll \beta \frac{dV}{dx}$ with suffices 1, 2.

It follows from (*b*) and eqns. (15) and (17) that $\cos\psi \simeq 0$ and so $\sin\psi \simeq \pm 1$. (If $\cos\psi$ differs from zero by a first-order small quantity, $|\sin\psi|$ differs from unity only in the second order of small quantities.)

Eqns. (14) and (16) are the only equations of significance, and these become

$$\frac{dV_1}{dx} = -\frac{\eta V_{30}}{2} \beta_1 V_2 \sin\psi \quad . \quad . \quad . \quad . \quad (19)$$

$$\frac{dV_2}{dx} = -\frac{\eta V_{30}}{2} \beta_2 V_1 \sin\psi \quad . \quad . \quad . \quad (20)$$

Eliminating V_2 between eqns. (19) and (20) gives

$$\frac{d^2 V_1}{dx^2} = \left(\frac{\eta V_{30}}{2}\right)^2 \beta_1 \beta_2 V_1 \quad . \quad . \quad . \quad . \quad (21)$$

remembering that $\sin\psi = \pm 1$. V_2 also satisfies this equation.

An appropriate solution, bearing in mind that the signal voltage will be finite at $x = 0$, but the idling voltage will be zero, is

$$V_1 = V_{10} \cosh \alpha_0 x \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$$V_2 = \mp V_{10} \sqrt{\left(\frac{\beta_2}{\beta_1}\right)} \sinh \alpha_0 x \quad . \quad . \quad . \quad (23)$$

where

$$\alpha_0 = \frac{\eta V_{30}}{2} \sqrt{(\beta_1 \beta_2)} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

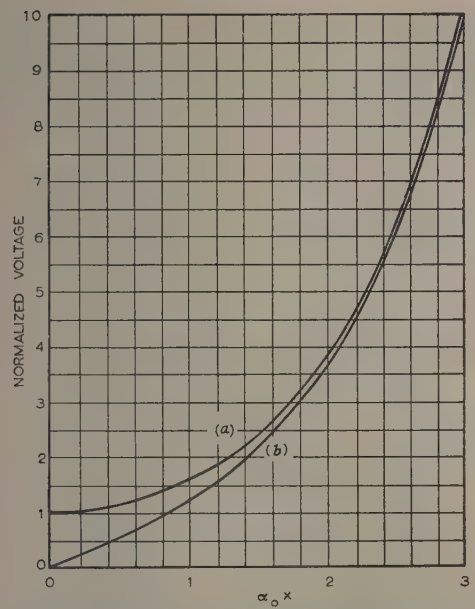


Fig. 1.—Growth of normalized signal- and idling-wave voltages with distance along the line.

(*a*) Signal wave, $\frac{V_1}{V_{10}}$.

(*b*) Idling wave, $\frac{V_2}{V_{10}}\sqrt{\frac{\omega_1}{\omega_2}}$.

The signal and idling wave amplitudes as given by eqns. (22) and (23) are plotted as functions of $\alpha_0 x$ in Fig. 1.

The useful gain of the amplifier is less than $\alpha_0 l$. It is in fact given by

$$\text{Gain} = \log_e \cosh \alpha_0 l \text{ nepers} \quad . \quad . \quad . \quad (25)$$

for a line of length l, or in decibels,

$$\text{Gain} = 20 \log_{10} \cosh \alpha_0 l \text{ decibels} \quad . \quad . \quad . \quad (26)$$

For a long line, $\alpha_0 l$ will be large, and $\cosh \alpha_0 l \simeq \frac{1}{2} \exp \alpha_0 l$. Hence, an approximate gain formula, valid only if the gain is large, is

$$\text{Gain} = (7{\cdot}686\alpha_0 l - 6) \text{ decibels} \quad . \quad . \quad . \quad (27)$$

This is reminiscent of the gain formula for a beam-type travelling-wave tube. In this case, however, only two waves are needed to satisfy the initial conditions, so the initial loss is only 6 dB, whilst in the travelling-wave tube three waves are involved, and the initial loss is $20 \log_{10} 3$, or 9·54 dB.

Let us now consider the bandwidth obtainable from a parametric travelling-wave amplifier. It follows from eqns. (24) and (2) that

$$\alpha_0 = \frac{\eta V_{30}}{2u}\sqrt{[\omega_1(\omega_3 - \omega_1)]} \quad . \quad . \quad . \quad (24a)$$

Using suitably normalized frequencies this can be written

$$\alpha_0^2 = \Omega_3\Omega_1 - \Omega_1^2 \quad . \quad . \quad . \quad . \quad . \quad (24b)$$

which is the equation of a circle, plotted in Fig. 2.

It should not be expected that this very wide-band performance will be realizable in practice with a periodically loaded line. Amongst other things, higher harmonics of the signal and idling frequencies and high-frequency intermodulation products may well lead to resonances in the short line sections separating the crystals at certain specific signal frequencies. These may lead

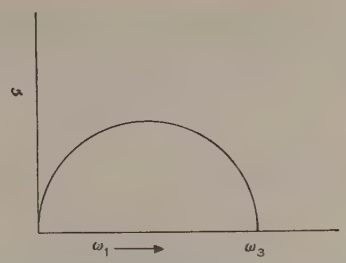


Fig. 2.—Gain as a function of signal frequency, ω_1, at a fixed pump frequency, ω_3.

$$\alpha_{max} = \frac{\eta V_{30}\omega_3}{4u}$$

to dips in the gain curve, and may also, more seriously, lead to instability. Experimental evidence is needed to establish the usable frequency range of an amplifier of this kind.*

Eqn. (24*b*) also shows that the dependence of gain on pumping frequency is parabolic, the signal frequency being fixed. This is shown in Fig. 3. Thus, increasing the pumping frequency will always increase the gain per unit length.

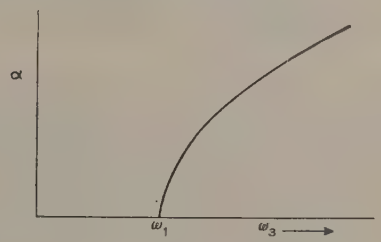


Fig. 3.—Gain as a function of pump frequency, ω_3, at a fixed signal frequency, ω_1.

(4) CONSERVATION OF POWER

It will now be shown that eqns. (14), (16) and (18) imply conservation of power. To the first approximation, that is to say $\alpha \ll \beta_1, \beta_2, \beta_3$, the characteristic impedance of the line is the same for all three waves and has the value

$$Z_0 = \sqrt{\frac{L_0}{C_0}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

This is easily proved from eqn. (4). If P_n is the power associated with the wave of voltage V_n we have

$$P_n = \frac{V_n^2}{2Z_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

with $n = 1, 2, 3$.

Since $$\frac{d(V_n^2)}{dx} = 2V_n\frac{dV_n}{dx}$$

it is a simple matter to express eqns. (14), (16) and (18) in terms of power instead of voltage, using eqn. (29). The result is

$$\frac{dP_1}{dx} = \mp \eta\beta_1 \sin\psi\,(2Z_0P_1P_2P_3)^{1/2} \quad . \quad . \quad . \quad (30)$$

$$\frac{dP_2}{dx} = \mp \eta\beta_2 \sin\psi(2Z_0P_1P_2P_3)^{1/2} \quad . \quad . \quad . \quad (31)$$

$$\frac{dP_3}{dx} = \pm \eta\beta_3 \sin\psi(2Z_0P_1P_2P_3)^{1/2} \quad . \quad . \quad . \quad (32)$$

* There is also an anomalous behaviour in the vicinity of synchronism, when $\omega_1 \simeq \frac{1}{2}\omega_3$. This question is not dealt with in the paper.

Using eqn. (8) it is easy to see that

$$\frac{dP_1}{dx} + \frac{dP_2}{dx} + \frac{dP_3}{dx} = 0$$

and hence

$$P_1 + P_2 + P_3 = \text{constant} \quad . \quad . \quad . \quad . \quad (33)$$

and so power is conserved.

(5) EFFECT OF LOSSES

So far, losses have been neglected. Consider now the practical form of parametric travelling-wave amplifier shown in Fig. 4.

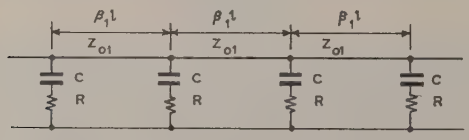


Fig. 4.—Diode-loaded transmission line.

This consists of a diode-loaded transmission line. The equivalent circuit of the diodes is a capacitance in series with a resistance. It will be assumed that

$$\beta l \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

This is not the optimum situation, for the gain per diode can be increased by increasing βl. The inequality (34) leads to a much simpler analysis, however, for it is then possible to smooth out the diode loads into an equivalent distributed load. In doing this we replace the equivalent series circuit of a diode by the corresponding parallel circuit. Assuming $(\omega CR)^2 \ll 1$ the equivalent capacitance, C_p, and conductance, G, of the parallel circuit are

$$C_p \simeq C \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

$$G \simeq \omega^2 C^2 R \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

Suppose the capacitance per unit length of the unloaded line is C_1 and the inductance per unit length is L_1. Then the effective primary constants of the loaded line at a frequency ω_n and at a very small voltage level are

Equivalent inductance per unit length $= L_1$

Equivalent capacitance per unit length $= C_1 + C/l$

Equivalent leakance per unit length $= \omega_n^2 C^2 R/l$

The secondary constants are

$$Z_0 \simeq \sqrt{\left(\frac{lL_1}{lC_1 + C}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

$$\beta = \omega_n \sqrt{\left[L_1\left(C_1 + \frac{C}{l}\right)\right]} \quad . \quad . \quad . \quad (38)$$

$$\alpha_n = \frac{\omega_n^2 C^2 R Z_0}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

It will be assumed that the value of α_n in eqn. (39), though deduced for a single low-voltage wave of frequency ω_n, will still hold at least in the first approximation in the presence of two waves of different frequencies.

It will also be assumed that the transfer of power from the pumping wave to the signal and idling waves is substantially unchanged by a small amount of resistance in series with the non-linear capacitance.

If these assumptions are tenable, the power gain per unit length given by eqn. (30) must be reduced by a power loss per unit length equal to $2\alpha_1 P_1$, and similar terms must be included in eqns. (31) and (32).

Hence we find

$$\frac{dP_1}{dx} = \mp \eta\beta_1 \sin\psi(2Z_0 P_1 P_2 P_3)^{1/2} - 2\alpha_1 P_1$$

$$\frac{dP_2}{dx} = \mp \eta\beta_2 \sin\psi(2Z_0 P_1 P_2 P_3)^{1/2} - 2\alpha_2 P_2$$

$$\frac{dP_3}{dx} = \pm \eta\beta_3 \sin\psi(2Z_0 P_1 P_2 P_3)^{1/2} - 2\alpha_3 P_3$$

In terms of voltages, these equations can be written

$$\frac{dV_1}{dx} = -\frac{\eta}{2}\beta_1 \sin\psi V_2 V_3 - \alpha_1 V_1 \quad . \quad . \quad (40)$$

$$\frac{dV_2}{dx} = -\frac{\eta}{2}\beta_2 \sin\psi V_1 V_3 - \alpha_2 V_2 \quad . \quad . \quad . \quad (41)$$

$$\frac{dV_3}{dx} = +\frac{\eta}{2}\beta_3 \sin\psi V_1 V_2 - \alpha_3 V_3 \quad . \quad . \quad . \quad (42)$$

It is assumed that the ohmic losses are small, and it is therefore permissible to neglect $\alpha_1 V_1$ and $\alpha_2 V_2$ in eqns. (40) and (41). We are also assuming small signals, and it is then possible to neglect the term $\frac{1}{2}\eta\beta_3 \sin\psi V_1 V_2$ in eqn. (42), in comparison with $\alpha_3 V_3$. More precisely, the following inequalities are assumed:

$$\left.\begin{aligned} \alpha_1 &\ll \frac{\eta}{2} V_3 \sin\psi\beta_1\left(\frac{V_2}{V_1}\right) \\ \alpha_2 &\ll \frac{\eta}{2} V_3 \sin\psi\beta_2\left(\frac{V_1}{V_2}\right) \\ \alpha_3 &\gg \frac{\eta V_3}{2} \sin\psi_3\left(\frac{V_1 V_2}{V_3^2}\right) \end{aligned}\right\} \quad . \quad . \quad . \quad (43)$$

These conditions are compatible if the losses are small and the signal voltage is always much less than the pumping voltage. Under these conditions, eqns. (40) and (42) become

$$\frac{dV_1}{dx} = -\frac{\eta V_3}{2}\beta_1 \sin\psi V_2 \quad . \quad . \quad . \quad . \quad (44)$$

$$\frac{dV_2}{dx} = -\frac{\eta V_3}{2}\beta_2 \sin\psi V_1 \quad . \quad . \quad . \quad . \quad (45)$$

$$\frac{dV_3}{dx} = -\alpha_3 V_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

It is easily verified that the appropriate solution (which is exact) is

$$V_1 = V_{10} \cosh \int \alpha dx \quad . \quad . \quad . \quad . \quad . \quad (47)$$

$$V_2 = -V_{10}\sqrt{\frac{\beta_2}{\beta_1}} \sinh \int \alpha dx \quad . \quad . \quad . \quad (48)$$

where

$$\alpha = \frac{\eta V_{30}}{2}\sqrt{(\beta_1\beta_2)}\varepsilon^{-\alpha_3 x} \quad . \quad . \quad . \quad . \quad (49)$$

and V_{10} and V_{30} are the values of V_1 and V_3 at $x = 0$. Substituting eqn. (49) in (47), the voltage amplitude of the signal wave is found to be

$$V_1 = V_{10} \cosh\left(\alpha_0 \frac{1 - \varepsilon^{-\alpha_3 x}}{\alpha_3}\right) \quad . \quad . \quad . \quad (50)$$

where $$\alpha_0 = \frac{\eta V_{30}}{2}\sqrt{(\beta_1\beta_2)} \quad . \quad . \quad . \quad . \quad (51)$$

In Fig. 5, eqn. (50) is plotted for several values of the ratio α_3/α_0. It will be observed that an increase of line length leads to an increase in gain up to a limiting value determined by the

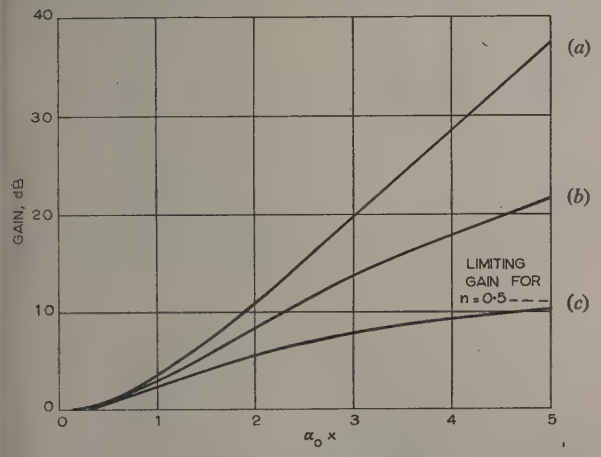


Fig. 5.—Gain as a function of line length with the ratio of pumping-wave attenuation coefficient (α_3) to initial growth constant (α_0) as a parameter.

(*a*) $\alpha_3 = 0$.
(*b*) $n = \alpha_3/\alpha_0 = 0{\cdot}2$.
(*c*) $n = \alpha_3/\alpha_0 = 0{\cdot}5$.

ratio α_3/α_0. This is also seen clearly from eqn. (50) in which, if $\alpha_3 x$ is very large, V_1 tends to the value

$$V_{1max} = V_{10}\cosh\frac{\alpha_0}{\alpha_3} \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Thus, the maximum gain obtainable is that which would be obtained in the absence of losses by a line of length $1/\alpha_3$, other things remaining unchanged.

The validity of eqn. (50) depends, it must be remembered, on the inequalities set out in (43). As distance along the line increases, V_3 will decrease relative to V_1 and V_2, partly because V_1 and V_2 are growing, but also because V_3 is decreasing owing to ohmic attenuation. The last inequality in (43) remains valid under these conditions, but ultimately the first two inequalities will be reversed, and the signal and idling waves must take the form

$$\left.\begin{aligned} V_1 &= A_1\varepsilon^{-\alpha_1 x} \\ V_2 &= A_2\varepsilon^{-\alpha_2 x} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (53)$$

Thus, the limiting gain given by eqn. (52) will not be maintained indefinitely as x increases, and indeed it will not quite be realized in practice. There will in fact be an optimum length of line for maximum gain. Without making a detailed study, it seems reasonable to expect that this optimum length will lie between the length at which the optimum gain is almost realized, which we take as $1/\alpha_3$ according to eqn. (50), and the length at which attenuation begins to become significant at the signal frequency in the absence of a pumping wave, which we take as $1/\alpha_1$. Hence, the optimum length of line is expected to be of the order $(\alpha_1\alpha_3)^{-1/2}$. The maximum gain will be less than that predicted by eqn. (52) by an amount which increases as α_1 increases.

(6) SATURATION EFFECTS

The foregoing Sections have all neglected the decrease in pumping-wave amplitude due to extraction of power from this wave to feed the growing signal and idling waves. This matter will now be considered, but to simplify the analysis, ohmic losses will be neglected.

We start with eqns. (14) and (16) together with the equation of conservation of power, which may be written

$$V_1^2 + V_2^2 + V_3^2 = V_{10}^2 + V_{30}^2 \quad . \quad . \quad . \quad (54)$$

Eqns. (14) and (16) are satisfied by the following expressions for V_1 and V_2:

$$V_1 = V_{10}\cosh\int \alpha dx \quad . \quad . \quad . \quad . \quad . \quad (55)$$

$$V_2 = -V_{10}\sqrt{\frac{\beta_2}{\beta_1}}\sinh\int \alpha dx \quad . \quad . \quad . \quad (56)$$

where $$\alpha = \frac{\eta V_3}{2}\sqrt{(\beta_1\beta_2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

From eqns. (54) and (57) it is easily shown that

$$\alpha^2 = \frac{\eta^2 V_{30}^2}{4}\beta_1\beta_2\left[1 + \frac{V_{10}^2}{V_{30}^2}\left(1 - \cosh^2\int \alpha dx - \frac{\beta_2}{\beta_1}\sinh^2\int \alpha dx\right)\right]$$

or $$\alpha^2 = \frac{\eta^2 V_{30}^2}{4}\beta_1\beta_2\left(1 - \frac{V_{10}^2}{V_{30}^2}\frac{\beta_3}{\beta_1}\sinh^2\int \alpha dx\right) \quad . \quad (58)$$

On physical grounds we know that the signal and idling waves cannot grow indefinitely, since there is only a finite amount of power available to feed them. The largest possible signal corresponds to the largest possible value of $\int \alpha dx$, and a maximum of $\int \alpha dx$ will be associated with the condition $\alpha = 0$. Putting $\alpha = 0$ in eqn. (58) we find

$$\left(\sinh^2\int \alpha dx\right)_{max} = \frac{V_{30}^2\beta_1}{V_{10}^2\beta_3} \quad . \quad . \quad . \quad (59)$$

Remembering that the signal power, P_1, is given by

$$P_1 = P_{10}\cosh^2\int \alpha dx \quad . \quad . \quad . \quad . \quad (60)$$

it follows from eqn. (59) that

$$P_{1max} - P_{10} = P_{30}\frac{\omega_1}{\omega_3}$$

or $$\frac{P_{1max} - P_{10}}{\omega_1} = \frac{P_{30}}{\omega_3}$$

In the same way, we find the value of P_{2max} to be

$$P_{2max} = P_{30}\frac{\omega_2}{\omega_3}$$

Combining these equations,

$$\frac{P_{1max} - P_{10}}{\omega_1} = \frac{P_2}{\omega_2} = \frac{P_{30}}{\omega_3} \quad . \quad . \quad . \quad . \quad (61)$$

Thus, at the position of maximum gain, the powers associated with the signal, idling and pumping waves satisfy the Manley–Rowe relations.[4] Note that $P_{20} = 0$, and that $P_3 = 0$ at the position of maximum gain. All the pumping power has been transferred to the signal and idling waves, the sharing-out being determined by the Manley–Rowe relations. The maximum power gain can easily be found explicitly from eqn. (61), and is

$$\frac{P_{1max}}{P_{10}} = 1 + \frac{\omega_1}{\omega_3}\frac{P_{30}}{P_{10}} \quad . \quad . \quad . \quad . \quad (61a)$$

Having found a maximum value for $\int \alpha dx$, the question of the behaviour of P_1 for further increase in x naturally arises. Also, we would like to know the value, l_1, of x at which this maximum occurs.

Differentiating eqn. (58) and simplifying, we get

$$\frac{d\alpha}{dx} = -\left(\frac{\eta}{2}\right)^2 \beta_2\beta_3 V_{10}^2 \tfrac{1}{2} \sinh 2\int \alpha dx \quad . \quad . \quad (62)$$

If we assume that the maximum gain is small, so that $\int \alpha dx$ is always small, we can replace the sinh term by its argument, and get

$$\left.\begin{aligned} \frac{d\alpha}{dx} &= -\left(\frac{\eta}{2}\right)^2 \beta_2\beta_3 V_{10}^2 \int \alpha dx \\ \text{or} \qquad \frac{d^2\alpha}{dx^2} &+ \left[\left(\frac{\eta}{2}\right)^2 \beta_2\beta_3 V_{10}^2\right]\alpha = 0 \end{aligned}\right\} \quad . \quad . \quad (63)$$

The appropriate solution is

$$\alpha = \alpha_0 \cos qx \quad . \quad . \quad . \quad . \quad . \quad (64)$$

with

$$\alpha_0 = \frac{\eta}{2}\sqrt{(\beta_1\beta_2)}V_{30} \quad . \quad . \quad . \quad . \quad (65)$$

$$q = \frac{\eta}{2}\sqrt{(\beta_2\beta_3)}V_{10} \quad . \quad . \quad . \quad . \quad (66)$$

It follows from eqns. (55) and (64) that

$$V_1 = V_{10} \cosh\left(\alpha_0 \frac{\sin qx}{q}\right) \quad . \quad . \quad . \quad (67)$$

When qx is sufficiently small, $\sin qx/q \to x$, and eqn. (67) reduces to the equation derived previously for the case in which V_3 was assumed constant. However, as x increases, V_1 varies periodically, having maxima of magnitude

$$\left.\begin{aligned} V_{1max} &= V_{10} \cosh\left(\frac{\alpha_0}{q}\right) \\ \text{or} \qquad V_{1max} &= V_{10} \cosh\left[\sqrt{\left(\frac{\beta_1}{\beta_3}\right)}\frac{V_{30}}{V_{10}}\right] \end{aligned}\right\} \quad . \quad . \quad (68)$$

The maxima occur at values of x given by

$$x_{max} = (2n+1)l_1 \quad . \quad . \quad . \quad . \quad . \quad (69)$$

where $l_1 = \dfrac{\pi}{2q} = \dfrac{\pi}{\eta V_{10}\sqrt{(\beta_2\beta_3)}}$

Thus, we see that the position of the first maximum, l_1, moves towards the input end of the line as the input signal increases. This behaviour is reminiscent of that of a klystron, in which the first maximum of r.f. current in the beam moves towards the bunching cavity as the drive power is increased.

Rather surprisingly, eqn. (69) predicts that l_1 increases with increasing signal frequency. This is because the idling frequency, ω_2, decreases as the signal frequency, ω_1, is increased. Hence β_2 decreases, and so, according to eqn. (69), l_1 increases. It is particularly to be noted that l_1 is independent of V_{30} under the assumed conditions of small maximum gain, depending only on the signal amplitude. For very small signal levels, l_1 becomes very large, and would greatly exceed the length of any practical amplifier, so that the periodic behaviour would then be unobservable. This is physically reasonable, for the amount of power extracted from the pumping wave will be very small under these conditions.

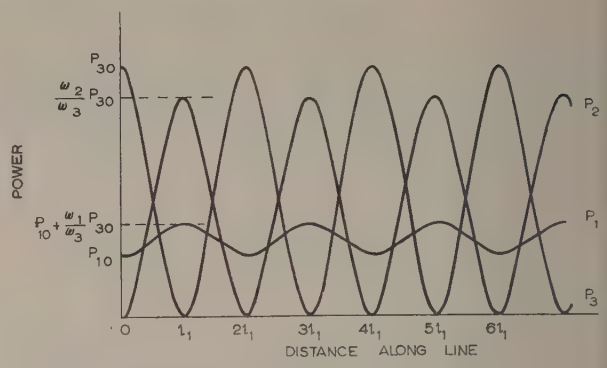


Fig. 6.—Periodic power transfer between signal and idling waves and pumping wave.

The general behaviour of the three waves, when periodic power transference occurs, is indicated by Fig. 6.

The approximations made in deriving eqn. (68) limit its application very severely, for the term $\sqrt{\left(\frac{\beta_1}{\beta_3}\right)}\frac{V_{30}}{V_{10}}$ must not exceed $\frac{1}{2}$ if the approximation employed in deriving eqn. (63) is to be reasonably satisfactory.

In most cases of practical interest, $V_{10} \ll V_{30}$ so that, in effect, eqn. (68) can only be employed when $\beta_1 \ll \beta_3$, i.e. when the signal frequency is much less than the pump frequency.

However, on physical grounds it seems likely that a periodic transference of power between the three waves will occur even when $\sqrt{\left(\frac{\beta_1}{\beta_3}\right)}\frac{V_{30}}{V_{10}}$ greatly exceeds unity, and we will now study this case.

To do so, let us write

$$\left.\begin{aligned} y &= 2\int \alpha dx \\ \frac{dy}{dx} &= 2\alpha \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (70)$$

Substituting these in eqn. (62) we find

$$\frac{d^2y}{dx^2} = -\left(\frac{\eta}{2}\right)^2 \beta_2\beta_3 V_{10}^2 \sinh y \quad . \quad . \quad . \quad (71)$$

There is a one-to-one correspondence between this equation and the equation for the vibration of a mass restrained by a spring having a non-linear restoring force proportional to the

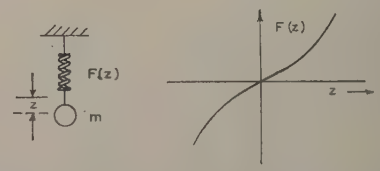


Fig. 7.—Mechanical analogy. Non-linear oscillatory system with restoring force proportional to hyperbolic sine of displacement.

hyperbolic sine of the displacement, as shown in Fig. 7. The corresponding equation is

$$\left.\begin{aligned} m\frac{d^2z}{dt^2} &= -F(z) \\ m\frac{d^2z}{dt^2} &= -K \sinh z \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (72)$$

It is physically obvious in this case that the motion will be periodic in time. It is also clear that the periodic time, τ, will decrease as K increases, and also, for a given K, τ will decrease as the amplitude of oscillation z_{max} increases. The last result follows because the 'average stiffness' of the spring increases as the amplitude of oscillation increases.

Applying these qualitative results to the parametric amplifier, we see that the length of line, l_1, at which maximum gain occurs will be less than that predicted by eqn. (69) as V_{10} increases. Furthermore, l_1 will also decrease as the maximum gain $\int_0^{l_1} \alpha dx$ increases, V_{10} being held constant. This is equivalent to saying that l_1 will decrease as V_{30} increases, with V_{10} constant.

It is also of interest to compare eqn. (58) with the energy conservation equation of the mechanical oscillatory system. The comparison is simpler if eqn. (58) is written in a slightly modified but equivalent form:

$$\frac{1}{2}\left(\frac{dy}{dx}\right)^2 + \left(\frac{\eta}{2}\right)^2 \beta_2\beta_3 V_{10}^2 \int_0^y \sinh y dy = \left(\frac{\eta}{2}\right)^2 \beta_1\beta_2 V_{30}^2 \quad (73)$$

$$\frac{1}{2}m\left(\frac{dz}{dt}\right)^2 + K\int_0^z \sinh z dz = W \quad . \quad . \quad . \quad (74)$$

Considering first the mechanical system, we see that the total energy, W, is shared between kinetic and potential energy. When the kinetic energy vanishes, the potential energy has its maximum value, W.

The corresponding formula in the amplifier case gives an upper limit to the gain, and this is, in fact, eqn. (61) which we have already derived.

A more useful result is obtained by noting that the maximum kinetic and potential energies are equal, and using this to get an approximate frequency of oscillation, assuming a sinusoidal waveform to relate amplitude and velocity. When the corresponding calculation is carried through for the parametric amplifier, it leads to the following result for the length of line at which the first gain maximum occurs:

$$l \simeq l_1 \left[\frac{\int_0^l \alpha dx}{\sinh \int_0^l \alpha dx} \right] \quad . \quad . \quad . \quad . \quad . \quad (75)$$

where l_1 is given by eqn. (69).

The factor in square brackets is always less than unity, so that the distance of the first maximum from the input decreases with increase of gain. This is shown in Fig. 8.

Finally, it is emphasized that the periodic transfer of power between signal and idling waves on the one hand and pumping wave on the other, discussed in this Section, occurs by quite a different mechanism from that discussed by Tien.[2] In Tien's paper, the periodicity arises from a deviation in the value of one of the phase coefficients, so that eqn. (8) is not satisfied.

In our case, it is assumed throughout that eqn. (8) is satisfied and the periodicity arises from the reduction in pumping power as the signal and idling waves grow, and from subsequent exchange of energy between these waves. In Tien's case, signals are assumed sufficiently small for the pumping wave to be substantially constant in amplitude.

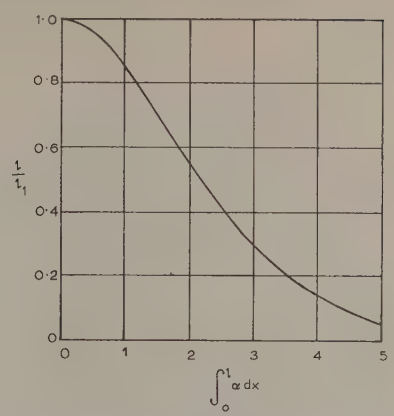


Fig. 8.—Position of first signal-wave maximum under large gain conditions, l, expressed in terms of the corresponding quantity, l_1, under small-gain conditions, and the maximum gain, $\int_0^l \alpha dx$.

(7) CONCLUSIONS

Some factors influencing the performance of travelling-wave parametric amplifiers have been discussed. In particular, the practical significance of ohmic losses in limiting the maximum gain obtainable is discussed, and a saturation phenomenon of a periodic character is predicted when the signal power is comparable with the pumping power.

The need for experimental evidence is stressed, and experiments are now in hand which are designed to test the theoretical results of the paper.

(8) ACKNOWLEDGMENTS

The author is grateful to his colleagues Mr. P. N. Robson and Mr. A. P. Jurkus, and to Mr. J. D. Pearson and Mr. D. Trevena of Ferranti Ltd., Wythenshawe, for very helpful discussions. The work described forms part of a research programme supported by C.V.D., to whom the author wishes to express acknowledgment.

(9) REFERENCES

(1) TIEN, P. K., and SUHL, H.: 'A Travelling-Wave Ferromagnetic Amplifier, *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 700.

(2) TIEN, P. K.: 'Parametric Amplification and Frequency Mixing in Propagating Circuits', *Journal of Applied Physics*, 1958, **29**, p. 1347.

(3) CULLEN, A. L.: 'A Travelling-Wave Parametric Amplifier', *Nature*, 1958, **181**, p. 332.

(4) MANLEY, J. M., and ROWE, H. E.: 'Some General Properties of Nonlinear Elements—Part I. General Energy Relations', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 904.

[The discussion on the above paper will be found on page 123.]

621.372.832.43.015.4

The Institution of Electrical Engineers
Paper No. 3031 E
Oct. 1959
©

SOME PROPERTIES OF TRAVELLING-WAVE RESONANCE

By J. R. G. TWISLETON, B.Sc., Associate Member.

(*The paper was first received 17th January, and in revised form 13th April, 1959. It was published in October, 1959, and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION *9th November, 1959.*)

SUMMARY

The paper gives a description of travelling-wave resonance in a waveguide ring fed through a directional coupler, and a resonance theory is developed which describes the behaviour of waves in the ring and the reflection and transmission in the feeder in the neighbourhood of resonance. The theory allows for equal propagation coefficients for the ring waves, or for a different propagation coefficient for one of the waves, and for a reflecting discontinuity in the ring. A method of investigating the ring resonance is described using a travelling probe, and experimental results are given for the case of equal ring-wave propagation coefficients, which substantiate the theory.

LIST OF SYMBOLS

L, L_1, L_2 = Length of ring and sections of ring.
$\Gamma = (\alpha + j\beta)L$ = Ring propagation coefficient; $\beta = \omega/v_p$.
ω, v_p = Angular frequency and phase velocity.
λ, λ_g = Wavelength and guide wavelength.
$\Gamma' + \Gamma$ = Total forward wave propagation coefficient, where $\Gamma' = (\alpha' + j\beta')L$.
r, t = Ring discontinuity voltage reflection and transmission coefficients.
r_1, a, b, c = Wave reflection and transmission coefficients for directional coupler; $k = c/a$.
C, D = Coupling factor and directivity.
ρ, τ = Reflection and transmission coefficients in waveguide feeder.
Q = Resonance Q-factor, where $q = n\pi/Q$, and n is an integer.
$s = 2n\pi S$ = Fractional deviation in resonant frequency from f_0, where $S = \delta f/f_0$.
$p = 2n\pi P$ = Fractional frequency deviation from mean, where $P = \Delta f f_0$.
z, y = Normalized impedance and admittance.

(1) INTRODUCTION

Increasing interest has arisen recently in the properties of waveguide ring circuits and microwave cavities which are excited through directional couplings, because of the phenomenon of travelling-wave resonance. In this introduction, the behaviour of a ring circuit and travelling-wave resonance is described briefly and applications are mentioned. Subsequently a resonance theory is developed for the two cases of equal and different propagation coefficients for oppositely travelling ring waves, and experimental verification of the theory is given for the first case.

In previous papers the effect of different coupling conditions and inequality of wave-propagation coefficients upon the circuit behaviour has been considered by Tischer.[1, 2] The basic analysis is similar to that given by Tischer, but the assumption regarding inequality of wave-propagation coefficients is different. In the present paper, the equations for the waves and wave ratios are developed in terms of the ring resonance parameters, Q-factor and frequency, to permit the derivation of the ring and coupler properties from experimental data. This representation also assists interpretation of the resonance behaviour.

(1.1) Description of Travelling-Wave Resonance

Ordinarily, waves launched in a ring by a non-directional coupler show a resonant build-up when the ring length is a whole multiple of the guide wavelength. At resonance, equal but oppositely directed waves combine to give a standing-wave pattern of field strength or voltage, whose position is fixed by the coupler. However, if waves are launched preferentially in one direction in the ring, for example by a directional coupler, the ring resonances may be characterized by the predominant build-up of one travelling wave. With directional coupling, only one travelling wave is launched in the ring. At resonance the wave reinforces itself after travelling round the ring, and in the absence of reflection the wave amplitude is constant at all points; this wave build-up is described as travelling-wave resonance.

The ring-wave amplitude may exceed that of the exciting wave, as in an ordinary cavity, and thus the circuit is useful for testing waveguide components at simulated high power with travelling-wave conditions.[3, 4] Other applications for the travelling-wave ring include the acceleration of electrons, and the idea of the feedback loop has been applied previously to a linear accelerator.[5] Alternatively, the directional excitation of resonant cavities is possible, and with a suitable field configuration such as that of H_{n10} oscillations in a cylindrical cavity, the continuous acceleration of electrons is possible. In these cases the use of a variable directional coupler may be advantageous, since it can be shown that the coupling factor has an optimum value depending upon the ring attenuation.[3] The ring circuit may also be useful for measurement of the properties of materials, for example ferrites, under travelling-wave conditions.

In some of these applications the ring propagation coefficients may be different for waves travelling in opposite directions. The difference of propagation coefficients in the accelerator is easily understood, since, in order to deliver energy to the electrons, a wave travelling with the electrons must be attenuated, whilst a wave travelling against the electrons may remain unmodified because there is no interaction. Similarly, with a ferrite device in the ring, phase shift or attenuation of one circuit wave may occur, in addition to the normal attenuation and phase shift. This non-reciprocity in the propagation can be allowed for in the theory of the ring circuit simply by assuming different propagation coefficients for oppositely travelling ring waves.

In the paper it is shown that the effect of unequal wave propagation coefficients is to modify the Q-factors, or frequencies, of the ring resonances. With reflection in the ring, or at the coupler, the characteristic voltage pattern is a standing wave, if the total ring-wave propagation coefficients are equal. Inequality of these coefficients can lead to travelling-wave conditions at resonance.

It may be noted that the effect of electron interaction with the circuit wave in a magnetron has been treated as a case of non-reciprocal propagation. In the analysis mentioned on page 1004

Mr. Twisleton is at the Research Laboratory of Associated Electrical Industries (Rugby) Ltd.

of Reference 6, it was assumed that currents induced at the anode segments were effective in modifying the propagation coefficient of one circuit wave. However, this representation may be hypothetical since later analyses for these devices[7, 8] show the existence of three circuit waves, rather than two waves as discussed here, which give rise to a different type of voltage pattern.

(2) ANALYSIS OF THE WAVEGUIDE RING CIRCUIT

The waveguide ring, coupled to an adjacent feeder by a directional coupler, is shown diagrammatically in Fig. 1. The

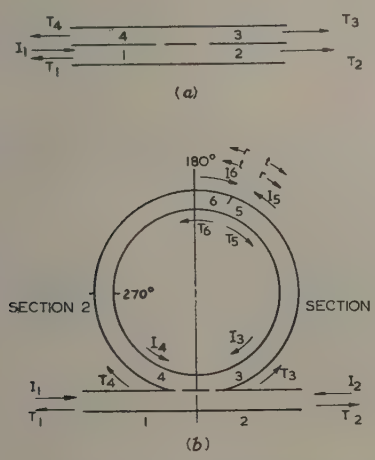


Fig. 1.—Waveguide ring circuit and coupled wave relations.

(*a*) Wave relations with matched terminations:

$T_1 = r_1 I_1$, $T_2 = bI_1$, $T_3 = aI_1$, $T_4 = cI_1$

(*b*) Wave relations in ring circuit:

$I_3 = T_5\varepsilon^{-\Gamma_1}$, $I_5 = T_3\varepsilon^{-\Gamma_1'-\Gamma_1}$

$I_4 = T_6\varepsilon^{-\Gamma_2'-\Gamma_2}$, $I_6 = T_4\varepsilon^{-\Gamma_2}$

wave pattern in the ring is regarded as a superposition of forward (anticlockwise) and reverse (clockwise) travelling waves. An incident wave, I_1, at the coupler input produces a forward travelling wave, T_3, in the ring, and a transmitted wave, T_2, at the output of the coupler. An electrical discontinuity and the coupler divide the ring into two sections. The discontinuity produces a reflected wave, $rT_3\varepsilon^{-\Gamma_1}$, in section 1 and a transmitted wave, $tT_3\varepsilon^{-\Gamma_1}$ in section 2, where $\Gamma_1 = (\alpha_1 + j\beta_1)L_1$ is the total propagation coefficient of section 1. In addition, a contribution to reverse waves arises from the imperfect directivity of the coupler.

The relation between the incident waves I_{1-4} and the transmitted waves T_{1-4} in the arms of the coupler are usually defined by $T = [S]I$, where $[S]$ is the scattering matrix. If the coupler is symmetrical, $[S]$ has the form

$$[S] = \begin{bmatrix} r_1 & b & a & c \\ b & r_1 & c & a \\ a & c & r_1 & b \\ c & a & b & r_1 \end{bmatrix} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In eqn. (1), a, b, and c are the wave transmission coefficients as defined in Fig. 1, and r_1 is the reflection coefficient in arm 1 when all the arms are terminated in matched loads. The coupling factor, C, and directivity, D, are usually defined as the power ratios $C = P_1/P_3$ and $D = P_3/P_4$, so that $|b| = \sqrt{(1 - 1/C)}$, $|a| = 1/\sqrt{C}$, and $|c| = \sqrt{[D(1 - 1/C)]}$. At the discontinuity in the ring, the reflected and transmitted waves are related by the matrix equation

$$\begin{bmatrix} T_5 \\ T_6 \end{bmatrix} = \begin{bmatrix} r & t \\ t & r \end{bmatrix} \begin{bmatrix} I_5 \\ I_6 \end{bmatrix} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

if the reflection is the same at each side of the discontinuity.

The system of equations defined by eqns. (1) and (2) may be solved to yield expressions for the waves in terms of the exciting wave, I_1, by inserting the relations between the incident and transmitted waves in the ring. In addition, an equation may be derived, putting $I_1 = 0$, specifying the resonant frequencies and damping of the ring resonances.

To allow for different propagation coefficients of the travelling waves, we assume that forward waves are modified by the factors $\varepsilon^{-\Gamma_1'}$ and $\varepsilon^{-\Gamma_2'}$, in addition to the normal factors $\varepsilon^{-\Gamma_1}$ and $\varepsilon^{-\Gamma_2}$, in sections 1 and 2 of the ring respectively. Assuming arm 2 to be terminated in a matched load, insertion of the boundary conditions $I_2 = 0$, $I_3 = T_5\varepsilon^{-\Gamma_1}$, $I_4 = T_6\varepsilon^{-\Gamma_2'-\Gamma_2}$, $I_5 = T_3\varepsilon^{-\Gamma_1'-\Gamma_1}$ and $I_6 = T_4\varepsilon^{-\Gamma_2}$ in eqns. (1) and (2) gives solutions for the waves and wave ratios as follows.

The reflection coefficient at the coupler input is

$$\rho = \frac{T_1}{I_1} = r_1 + a\left[\left(\frac{T_5}{T_6}\right)\varepsilon^{-\Gamma_1} + k\varepsilon^{-\Gamma_2}\right]\frac{T_6}{I_1} \quad . \quad . \quad (3)$$

and the transmission coefficient in the waveguide is

$$\tau = \frac{T_2}{I_1} = b + a\left[\left(\frac{T_5}{T_6}\right)k\varepsilon^{-\Gamma_1} + \varepsilon^{-\Gamma_2'-\Gamma_2}\right]\frac{T_6}{I_1} \quad . \quad . \quad (4)$$

where $\Gamma' = \Gamma_1' + \Gamma_2'$, and $k = c/a$ specifies the directivity of the coupler, i.e. $k = 0$ for directional coupling, $1 > k > 0$, for semi-directional coupling and $k = 1$ for non-directional coupling. From eqn. (2), the ratio of the ring waves at the discontinuity is

$$\frac{T_5}{T_6} = \frac{r\varepsilon^{-\Gamma_1'-\Gamma_1} + \left(\frac{T_4}{T_3}\right)t\varepsilon^{-\Gamma_2}}{t\varepsilon^{-\Gamma_1'-\Gamma_1} + \left(\frac{T_4}{T_3}\right)r\varepsilon^{-\Gamma_2}} \quad . \quad . \quad . \quad . \quad (5)$$

where $$T_3 = a\frac{Bk + C}{AC - BD}I_1, \; T_4 = a\frac{Ak + D}{AC - BD}I_1 \quad . \quad . \quad (6)$$

$$\frac{T_4}{T_3} = \frac{Ak + D}{Bk + C}, \text{ and}$$

$$\left.\begin{aligned} A &= 1 - rr_1\varepsilon^{-\Gamma_1'-2\Gamma_1} - b\varepsilon^{-\Gamma'-\Gamma} \\ B &= r_1 t\varepsilon^{-\Gamma} + br\varepsilon^{-\Gamma_2'-2\Gamma_2} \\ C &= 1 - bt\varepsilon^{-\Gamma} - rr_1\varepsilon^{-\Gamma_2'-2\Gamma_2} \\ D &= br\varepsilon^{-\Gamma_1'-2\Gamma_1} + r_1 t\varepsilon^{-\Gamma'-\Gamma} \end{aligned}\right\} \quad . \quad . \quad (7)$$

The ratio T_6/I_1 in eqns. (3) and (4) is

$$\frac{T_6}{I_1} = a\frac{[(Bk + C)t\varepsilon^{-\Gamma_1'-\Gamma_1} + (Ak + D)r\varepsilon^{-\Gamma_2}]}{AC - BD} \quad . \quad (8)$$

The frequencies and the damping of the ring resonances are found from the solution for ω of the equation

$$AC - BD = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Some solutions are discussed in Section 3.

(2.1) Ring Waves and the Waveguide Reflection and Transmission Coefficients with Directional Coupling

To simplify the equations of the previous Section it is convenient to consider special cases, and first we assume that the

coupler is directional and non-reflecting. The equations for the waves and wave ratios which are obtained with reflection at the coupler, but no reflection in the ring, show that the theory for this case is closely parallel to that given below. Putting $k = 0$ and $r_1 = 0$, the reflection and transmission coefficients in the main waveguide are

$$\rho = \frac{a^2 r \varepsilon^{-\Gamma}\{-2\Gamma_1+2\Gamma}}{(\varepsilon^{\Gamma} - bt\varepsilon^{-\Gamma'})(\varepsilon^{\Gamma} - bt) - b^2 r^2 \varepsilon^{-\Gamma'}} \quad . \quad . \quad (10)$$

$$\tau = b + \frac{a^2 \varepsilon^{-\Gamma'}[t\varepsilon^{\Gamma} - b(t^2 - r^2)]}{(\varepsilon^{\Gamma} - bt\varepsilon^{-\Gamma'})(\varepsilon^{\Gamma} - bt) - b^2 r^2 \varepsilon^{-\Gamma'}} \quad . \quad (11)$$

and the ring wave, T_3, is

$$T_3 = \frac{a\varepsilon^{\Gamma}(\varepsilon^{\Gamma} - bt)}{(\varepsilon^{\Gamma} - bt\varepsilon^{-\Gamma'})(\varepsilon^{\Gamma} - bt) - b^2 r^2 \varepsilon^{-\Gamma'}} \quad . \quad . \quad (12)$$

The ring wave ratios are

$$\left(\frac{T_5}{T_3}\right) = \frac{r\varepsilon^{-\Gamma_1-\Gamma_1}}{1 - bt\varepsilon^{-\Gamma}}, \quad \left(\frac{T_5}{T_3}\right)_5 = \frac{r\varepsilon^{\Gamma}}{\varepsilon^{\Gamma} - bt} \quad . \quad . \quad (13)$$

$$\left(\frac{T_4}{T_6}\right) = \frac{br\varepsilon^{-\Gamma_1}}{t - b(t^2 - r^2)\varepsilon^{-\Gamma}}, \quad \left(\frac{T_4}{T_6}\right)_6 = \frac{br}{t\varepsilon^{\Gamma} - b(t^2 - r^2)} \quad . \quad (14)$$

where $(T_5/T_3)_5$ and $(T_4/T_6)_6$ are the wave ratios at terminals 5 and 6 respectively.

(2.1.1) No Reflection in the Ring.

With no reflection in the ring, i.e. $r = 0$ and $t = 1$, the input reflection coefficient is zero and the transmission coefficient is

$$\tau = b + \frac{a^2 \varepsilon^{-\Gamma'-\Gamma}}{1 - b\varepsilon^{-\Gamma'-\Gamma}} \quad . \quad . \quad . \quad . \quad (15)$$

and the ring wave is

$$T_3 = \frac{aI_1}{1 - b\varepsilon^{-\Gamma'-\Gamma}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

with $T_4 = 0$. Since T_4 is zero, a purely travelling wave, T_3, exists in the ring. The ring wave T_3, and τ, exhibit a resonant behaviour when the imaginary part of $\Gamma' + \Gamma$ passes through zero. This occurs when $L = n\lambda g$, where n is an integer equal to the number of wave repeats round the ring, and T_3 attains a value at resonance given by

$$\left.\begin{aligned} T_3 &= \frac{aI_1}{1 - b\varepsilon^{-(\alpha'+\alpha)L}} \\ \text{and} \qquad \tau &= b + \frac{a^2 \varepsilon^{-(\alpha'+\alpha)L}}{1 - b\varepsilon^{-(\alpha'+\alpha)L}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (17)$$

For a given ring attenuation, the wave amplitude in the ring depends upon the coupling between the guide and ring, and it is of interest to note that, by differentiation of the expression for T_3, maximum transfer is obtained when the coupling factor, C, is given by

$$C_{opt} = \frac{1}{1 - \varepsilon^{-2(\alpha'+\alpha)L}} \quad . \quad . \quad . \quad . \quad (18)$$

Then $\tau = 0$.

The existence of an optimum value for the coupling can be shown as follows. At resonance, a wave T_3 launched in the ring arrives back at the starting point in phase with the entering wave. The transmitted wave in the waveguide aT_3 due to the ring wave T_3 is opposite in phase* to that transmitted through the coupler, i.e. bI_1, and hence cancellation occurs when the magnitudes of these components are equal. The magnification of T_3 is determined by the loaded Q-factor of the ring circuit which is a function of the coupling and the internal attenuation. The amplitudes of the two waves, T_3 and bI_1, vary inversely with the coupling factor b, and hence τ is zero at resonance for a particular value of the coupling.

In some applications, for example the testing of waveguide components and the accelerator, the power gain which can be obtained is important. Referring to Appendix 10.2 we find that, if the ring attenuation is small, the power gain with the optimum coupling is $Q_0/2n\pi$. The equation for T_3/I_1 given there, expressed in terms of Q-factor and frequency, shows that gain is obtained at the expense of bandwidth.

The case of large attenuation, and the use of another form of ring circuit for high-power testing, is discussed in Reference 3.

(2.2) Ring-Wave Behaviour with Reflection in the Ring

If reflection occurs at some point in the ring, a standing wave, rather than a travelling wave, is set up* as a result of successive wave transits and reflections. In addition, each ring-mode resonance, corresponding to $n\lambda g = L$, may be split into two closely spaced resonances. Experimentally we find that as the frequency is varied the ring waves and the reflected and transmitted waves in the main guide exhibit a double resonant peak.

The denominator of eqns. (10), (11), and (12) may be written as the product $(\varepsilon^{\Gamma} - \varepsilon^{\Gamma_1})(\varepsilon^{\Gamma} - \varepsilon^{\Gamma_2})$, where Γ_1 and Γ_2 are solutions of

$$(\varepsilon^{\Gamma} - bt\varepsilon^{-\Gamma'})(\varepsilon^{\Gamma} - bt) - b^2 r^2 \varepsilon^{-\Gamma'} = 0 \quad . \quad . \quad (19)$$

which are

$$\varepsilon^{\Gamma_{1,2}} = b\varepsilon^{-\Gamma'/2}\left\{t\cosh\frac{\Gamma'}{2} \pm \sqrt{\left[t^2\cosh^2\frac{\Gamma'}{2} - (t^2 - r^2)\right]}\right\} . \quad (20)$$

The reflection coefficient, ρ, and the waves T_3 and T_4 attain maximum values at frequencies near the resonant frequencies given by Γ_1 and Γ_2. The amplitude/frequency dependence of the ring waves has been discussed by Tischer[2] in terms of a vector diagram using a plot in the complex plane of the 'poles' $\varepsilon^{\Gamma_{1,2}}$ and the zeros [for example $\varepsilon^{\Gamma_3} = bt$ for eqn. (12)] of the wave functions. Here, as discussed in Section 4, a vector diagram representing the wave ratios T_5/T_3 and T_4/T_6 is used to describe the resonance behaviour of these ratios.

(3) THE Q-FACTORS AND FREQUENCIES OF THE RING RESONANCES

In an experimental investigation of ring-guide properties, the quantities measured most easily are the input-voltage reflection coefficient in the waveguide, and the standing-wave ratio in the ring, from which the frequencies and Q-factors of the resonances may be found. It is possible to express the wave functions in terms of these quantities. Assuming equal propagation coefficients for the forward and reverse waves, i.e. $\Gamma' = 0$, the resonance equation (20) becomes

$$\varepsilon^{\Gamma_{1,2}} = b(t \pm r) \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Eqn. (21) has solutions for $\omega = \omega_a + j\omega_b$, where $\omega_b = \omega_a/2Q$. Q being the loaded Q-factor, as follows:

$$\frac{\omega_a L}{v_p} = 2n\pi + \arg\left[b(t \pm r)\right] \quad . \quad . \quad . \quad . \quad (22)$$

* For a directional coupler the coupling factor a in eqn. (1) becomes ja.

* The possibility of reflection cancellation by a second discontinuity exists, leading to a travelling wave in part of the ring.

and $$\frac{\omega_b L}{v_p} = \frac{\omega_a L}{v_p} \frac{1}{2Q_{1,2}} = \alpha L - \log_\varepsilon |b(t \pm r)| \quad . \quad . \quad (23)$$

If arg $b(t \pm r)$ is zero, then since $\omega_a/v_p = 2\pi/\lambda_g$, the resonant frequency given by eqn. (22) is defined by $n\lambda_g = L$, where λ_g is the ring-guide wavelength at resonance and n is an integer equal to the number of wave repeats in the ring, as mentioned previously.

Now assuming arg $[b(t \pm r)]$ to be small, the change in resonant frequencies $f_{1,2}$ from f_0 are given by

$$2n\pi \frac{\delta f_{1,2}}{f_0} = \arg [b(t \pm r)] \quad . \quad . \quad . \quad . \quad (24)$$

and putting $\omega_a L/v_p = 2n\pi$, the Q-factors are

$$\frac{n\pi}{Q_{1,2}} = \alpha L - \log_\varepsilon |b(t \pm r)| \quad . \quad . \quad . \quad . \quad (25)$$

It is evident from eqn. (25) that we may define an internal Q-factor, Q_0, as

$$Q_0 = \frac{n\pi}{\alpha L} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

and a Q-factor, Q_c, arising from the external coupling as

$$Q_c = \frac{n\pi}{\log_\varepsilon \left|\frac{1}{b}\right|} = \frac{2n\pi}{\log_\varepsilon \left(\frac{C}{C-1}\right)} \quad . \quad . \quad . \quad (27)$$

introducing the coupling factor, C. Thus, the loaded Q of the resonances is

$$\frac{1}{Q_{1,2}} = \frac{1}{Q_0} + \frac{1}{Q_c} + n\pi \log_\varepsilon \left|\frac{1}{t \pm r}\right| \quad . \quad . \quad . \quad (28)$$

We see that, if $|t \pm r|$ is unity, the electrical discontinuity does not contribute to the loading of the resonance. Thus the condition $|t \pm r| = 1$ corresponds to a lossless discontinuity.

In accordance with eqns. (24) and (28), the resonance equation (21) may be written:

$$\left.\begin{aligned} \varepsilon^{n\pi\left(\frac{1}{Q_0}+\frac{1}{Q_c}-\frac{1}{Q_1}+2jS_1\right)} &= t + r \\ \varepsilon^{n\pi\left(\frac{1}{Q_0}+\frac{1}{Q_c}-\frac{1}{Q_2}+2jS_2\right)} &= t - r \end{aligned}\right\} \quad . \quad . \quad . \quad (29)$$

Here, $S_1 = \delta f_1/f_0$, and $S_2 = \delta f_2/f_0$ are the fractional frequency deviations of the resonances 1 and 2 from f_0. We note that $\frac{1}{2}(S_1 + S_2)$ is the mean fractional frequency deviation of the two resonances, and $S = (S_1 - S_2)$ is the fractional frequency separation of the resonances.

Referring to Appendix 10.1, expressions for r and t obtained from eqn. (29) are

$$r = \varepsilon^{q_0+q_e-\frac{q_1+q_2}{2}+j\frac{s_1+s_2}{2}} \sinh \frac{q_2 - q_1 + js}{2}$$

and $$t = \varepsilon^{q_0+q_e-\frac{q_1+q_2}{2}+j\frac{s_1+s_2}{2}} \cosh \frac{q_2 - q_1 + js}{2}$$

where $q_0 = n\pi/Q_0$, and $s_1 = 2n\pi S_1$, etc. Using the same notation,

$$\Gamma = q_0 + j\left(p + \frac{s_1 + s_2}{2}\right)$$

where $p = 2n\pi P$ and $P = \Delta f/f_0$, Δf being the frequency deviation from the mean frequency of the two resonances.

Equations for the wave ratios in terms of the frequencies and Q-factors of the ring resonances can now be obtained substituting for r, t, and Γ as given above, in eqns. (13) and (14). Referring to Appendix 10.1 we find that approximate expressions for ring-wave ratios are:

$$\left.\begin{aligned} \left(\frac{T_5}{T_3}\right)_5 &= \frac{\frac{q_2 - q_1}{2} + j\frac{S}{2}}{\frac{q_1 + q_2}{2} + jp} \varepsilon^E \\ \text{and} \quad \left(\frac{T_4}{T_6}\right)_6 &= \left(\frac{T_5}{T_3}\right)_5 \varepsilon^{-q_e} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (30)$$

Eqns. (30) show that the magnitude of the wave ratio is different in each section of the ring. This is unimportant in the analysis of the wave ratio/frequency curve, which has the familiar form of a resonance curve. In terms of actual Q-factors, frequency and frequency separation, the ring-wave ratio $(T_5/T_3)_5$, omitting the factor ε^E, is

$$\left(\frac{T_5}{T_3}\right)_5 = \frac{\frac{1}{2}\left(\frac{1}{Q_2} - \frac{1}{Q_1}\right) + jS}{\frac{1}{2}\left(\frac{1}{Q_1} + \frac{1}{Q_2}\right) + j2P} \quad . \quad . \quad . \quad (31)$$

When Q_1 is equal to Q_2, these factors and the fractional frequency separation, S, can easily be found from an experimental plot of ring-wave ratio against frequency, or wavelength. The derivation of these factors from such a plot, and special features of the ring-wave ratio, are discussed in the following Section.

(4) THE DERIVATION OF RESONANCE PARAMETERS

If the ring discontinuity is lossless, the loaded Q-factors of the two resonances must be equal. Eqn. (31) for the ring-wave ratio then simplifies to

$$\left.\begin{aligned} \left(\frac{T_5}{T_3}\right)_5 &= \frac{jS}{\frac{1}{Q} + j2P} \\ \text{and the phase angle of } (T_5/T_3)_5 \text{ is} \quad & \\ \frac{\pi}{2} - \arctan 2PQ & \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (32)$$

The wave ratio has a maximum value of SQ at the mean frequency of the two resonances, corresponding to $P = 0$, and at frequencies corresponding to $P = \pm S/2$ the ratio is $SQ/\sqrt{[1 + (SQ)^2]}$. Thus, the wave ratio only exceeds unity if $S > 1/Q$. Although the ring-wave ratio attains a maximum value at the mean resonant frequency, the ring-wave amplitudes are greatest at frequencies near to the resonant frequencies, corresponding to $P = \pm S/2$.

Both S and Q can be found from an experimental plot of ring-wave ratio against frequency or wavelength. Referring to Fig. 2, the loaded Q-factor is determined by the bandwidth of the curve at 0·707 of the maximum value and S can be found from the maximum value, or from the bandwidth at unit wave-ratio.

Thus, $1/Q = 2P'$, and $S = R_m \times 2P'$, where $2P'$ is the bandwidth at 0·707 of the maximum wave-ratio, R_m. Alternatively if $2P''$ is the bandwidth at unit wave-ratio, then $S^2 = (2P')^2 + (2P'')^2$.

If the ring discontinuity is lossy, Q_1 and Q_2 are not equal, and to evaluate these factors it is necessary to obtain more data than are provided by the curve of ring-wave ratio. The derivation of the resonance parameters in this case is discussed in Appendix 10.3.

The resonance behaviour of the wave ratio $(T_5/T_3)_5 = r/(1 - bt\varepsilon^{-\Gamma})$ can be clearly understood from the vector diagram

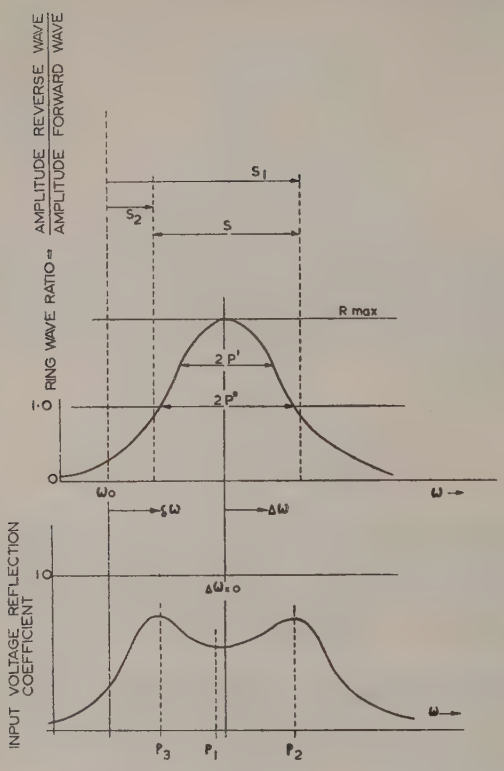


Fig. 2.—Frequency variation of ring-wave ratio and input reflection.

$P = \Delta\omega/\omega_0 \qquad p = 2n\pi P$

$S = \delta\omega/\omega_0 \qquad s = 2n\pi S$

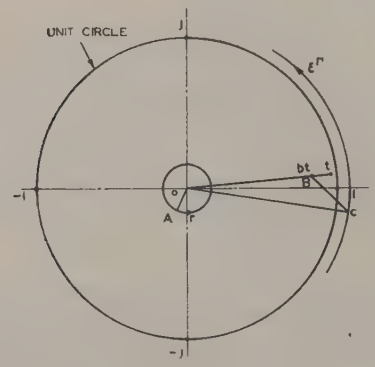


Fig. 3.—Vector diagram for ring-wave ratio with directional coupler.

$$\text{Ring-wave ratio} = \frac{OA}{BC} = \frac{|r\varepsilon^{\Gamma}|}{|\varepsilon^{\Gamma} - bt|}$$

For lossless discontinuity, $r = -j \sin \frac{1}{2}s\varepsilon^{j\frac{1}{2}s}$, $t = \cos \frac{1}{2}s\varepsilon^{j\frac{1}{2}s}$

of Fig. 3. The magnitude of the wave ratio is the ratio of the vectors $OA = r\varepsilon^{\Gamma}$ and $BC = \varepsilon^{\Gamma} - bt$. The vectors $r\varepsilon^{-\Gamma}$ and $\varepsilon^{-\Gamma}$ rotate in a clockwise direction with increasing frequency; it can be seen that, if $|r| > |1 - bt|$, which corresponds to $S > 1/Q$, the wave ratio passes through unity at two separate frequencies, and attains a maximum at an intermediate frequency.

The case of semi-directional coupling is of practical importance, and the variation in the wave ratio is perhaps best understood in terms of a vector diagram as just described. This is discussed in Appendix 10.4.

(5) RESONANCE BEHAVIOUR WITH UNEQUAL RING-WAVE PROPAGATION COEFFICIENTS

(5.1) The Q-Factors and Resonant Frequencies

So far in the discussion we have assumed equal total propagation coefficients for the forward and reverse waves. We now consider the changes in Q-factors and resonant frequencies which result from unequal total propagation coefficients.

It is convenient to write the resonance equation (20) in the form

$$\varepsilon^{\Gamma_{1,2}} = b\sqrt{(t^2 - r^2)}\varepsilon^{-\frac{1}{2}\Gamma' \pm \phi} \quad . \quad . \quad . \quad . \quad (33)$$

where $$\phi = \text{arc cosh}\left[\frac{t}{\sqrt{(t^2 - r^2)}} \cosh \frac{\Gamma'}{2}\right]$$

An approximate expression for ϕ may be obtained if $\Gamma'/2$ is assumed to be small and if $t/\sqrt{(t^2 - r^2)}$ is nearly unity. From Appendix 10.1, the ratio $t/\sqrt{(t^2 - r^2)}$ is found to be $\cosh \frac{1}{2}(q_2 - q_1 + js)$, so that omitting terms of higher power than the second in the series expansions for this quantity and for $\cosh \Gamma'/2$ we find that

$$\phi = \text{arc cosh}\left[1 + \frac{1}{2}\left(\frac{q_2 - q_1}{2} + j\frac{s}{2}\right)^2 + \frac{1}{2}\left(\frac{\Gamma'}{2}\right)^2\right]$$

With further approximation,

$$\phi = \tfrac{1}{2}\sqrt{[(q_2 - q_1 + js)^2 + \Gamma'^2]} \quad . \quad . \quad . \quad (34)$$

Using the expressions for r and t given in Appendix 10.1, we find from eqn. (33) that

$$\Gamma_{1,2} = -\frac{\Gamma'}{2} + q_0 - \frac{q_1 + q_2}{2} + j\frac{s_1 + s_2}{2} \pm \phi \quad . \quad (35)$$

Putting $\Gamma' = \alpha'' + j\beta''$, where α'' and β'' are the total changes in attenuation and phase shift of the forward travelling wave, the Q-factors, $\bar{q}_{1,2}$, and frequency deviation, $\bar{s}_{1,2}$, of the two resonances are found from eqn. (35) as follows:

$$\left.\begin{aligned} \bar{q}_{1,2} &= \frac{\alpha''}{2} + \frac{q_1 + q_2}{2} \pm \mathscr{R}\phi \\ \bar{s}_{1,2} &= -\frac{\beta''}{2} + \frac{s_1 + s_2}{2} \pm \mathscr{I}\phi \end{aligned}\right\} \quad . \quad . \quad . \quad (36)$$

and the frequency separation of the resonances is

$$\bar{s} = \mathscr{I}\sqrt{[(q_2 - q_1 + js)^2 + \Gamma'^2]}$$

First we consider how an inequality in ring-wave propagation coefficients influences the resonance Q-factors and frequencies.

In the simplest case, that of no reflection in the ring, the loaded Q-factors of the two resonances given by eqn. (36) are $q_0 + q_c + \alpha''$ and $q_0 + q_c$. It is evident that these factors correspond to forward- and reverse-wave resonance respectively. The two resonances occur at the same frequency if β'' is zero, or otherwise at different frequencies.

With reflection in the ring, the characteristic wave pattern is a standing wave which, as shown in Appendix 10.5, is so oriented that the field couples, or does not couple, with the discontinuity. It may be shown that different forward- and reverse-wave propagation coefficients give rise to unequal wave amplitudes at resonance, and also to rotation of the wave pattern. With a lossy discontinuity, travelling-wave conditions lead to equality

of the resonance Q-factors, and, alternatively, changes in Q-factors can arise from rotation of the standing-wave pattern.*

(5.2) The Ring-Wave Ratio, Waveguide Reflection and Transmission Coefficients

Eqns. (10), (11) and (12) for ρ, τ, and T_3 may be put in terms of the q- and s-parameters and $\Gamma_{1,2}$ given by eqn. (35), as follows:

$$\rho = a^2\left[\frac{\frac{q_2-q_1}{2}+j\frac{s}{2}}{\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}+\phi+jp\right)\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}-\phi+jp\right)}\right]\varepsilon^{-\Gamma_1-2\Gamma_1+2\Gamma+E} \quad . \quad (37)$$

$$\tau = b + a^2\left[\frac{\frac{q_1+q_2}{2}+jp}{\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}+\phi+jp\right)\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}-\phi+jp\right)}\right]\varepsilon^{-\Gamma'+E} \quad . \quad (38)$$

$$\frac{T_3}{T_1} = a\left[\frac{\frac{q_1+q_2}{2}+jp}{\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}+\phi+jp\right)\left(\frac{q_1+q_2}{2}+\frac{\Gamma'}{2}-\phi+jp\right)}\right]\varepsilon^{\Gamma} \quad . \quad (39)$$

where $b = \varepsilon^{-q_c}$, $a^2 = -2q_c$, and E is defined in Appendix 10.1. The ring-wave ratio at resonance with external excitation is obtained by substituting $p = -\beta''/2 \pm \mathscr{I}\phi$ in eqns. (30). We find that

$$\left(\frac{T_5}{T_3}\right)_5 = \left[\frac{\frac{q_2-q_1}{2}+j\frac{s}{2}}{\frac{q_1+q_2}{2}+j\left(-\frac{\beta''}{2}\pm\mathscr{I}\phi\right)}\right]\varepsilon^{E} \quad . \quad (40)$$

Referring to eqns. (37), (38), (39) and (40), the variation in ρ, T_3, etc., near resonance is determined by the bracketed quantities, rather than by the exponential factors, which produce only small changes in magnitude and phase.

In the general case of unequal propagation coefficients, the resonance behaviour is complex and we restrict the discussion to that of attenuation, or phase shift, of forward-travelling waves.

(5.3) Attenuation of Forward-Travelling Waves

To simplify matters, we further assume that the ring discontinuity is lossless. This implies that $q_1 = q_2 = q_0 + q_c$ and hence, from eqn. (34), $\phi = \frac{1}{2}\sqrt{(\alpha''^2 - s^2)}$. There are two cases to consider, α'' greater or less than s. If α'' is less than s, $\phi = \frac{1}{2}j\sqrt{(s^2 - \alpha''^2)}$. From eqn. (36), the Q-factors, frequencies and frequency separation of the pair of resonances are given by

$$\left.\begin{aligned}\bar{q}_{1,2} &= \frac{\alpha''}{2} + q_0 + q_c\\ \bar{s}_{1,2} &= \frac{s_1+s_2}{2} \pm \frac{1}{2}\sqrt{(s^2-\alpha''^2)}\\ \bar{s} &= \sqrt{(s^2-\alpha''^2)}\end{aligned}\right\} \quad . \quad . \quad . \quad (41)$$

The equations for $\bar{q}$ and $\bar{s}$ show that the Q-factors and frequency separation of the resonances are reduced by non-reciprocal attenuation, the mean frequency of the resonances remaining unaltered.

With increased attenuation corresponding to α'' greater than s, $\phi = \frac{1}{2}\sqrt{(\alpha''^2 - s^2)}$, the frequency separation of the resonances remains zero but the $\bar{q}$-factors become

$$\bar{q}_{1,2} = \frac{\alpha''}{2} + \frac{q_1+q_2}{2} \pm \frac{1}{2}\sqrt{(\alpha''^2 - s^2)} \quad . \quad . \quad (42)$$

Referring to eqns. (37), (38) and (39), we find that, for $\alpha'' < s$, the reflection, transmission, and wave amplitude, T_3, exhibit double maxima in the neighbourhood of resonance, whilst for $\alpha'' > s$ only a single maximum occurs. With the latter conditions, the ring-wave ratio at resonance, from eqn. (40), has its maximum value of

$$\left(\frac{T_5}{T_3}\right)_5 = \frac{\frac{q_2-q_1}{2}+j\frac{s}{2}}{\frac{q_1+q_2}{2}}$$

(5.4) Phase Shift of Forward-Travelling Waves

In this case $\Gamma' = j\beta''$ and when $q_1 = q_2$, $\phi = \frac{1}{2}j\sqrt{(s^2 + \beta''^2)}$. The $\bar{q}$-factors and resonant frequencies are given by

$$\left.\begin{aligned}&\bar{q}_{1,2} = q_0 + q_c\\ \text{and}\quad &\bar{s}_{1,2} = -\frac{\beta''}{2} + s_2 + \frac{1}{2}[s \pm \sqrt{(s^2+\beta''^2)}]\\ \text{with}\quad &\bar{s} = \sqrt{(s^2+\beta''^2)}\end{aligned}\right\} \quad . \quad (43)$$

The effect of phase shift is to increase the frequency separation of the resonances. If β'' is appreciably greater than s, the frequencies are given approximately by $\bar{s}_1 = \frac{1}{2}(s_1 + s_2)$ and $\bar{s}_2 = \frac{1}{2}(s_1 + s_2) - \beta''$.

With external excitation at these frequencies, the ring-wave ratios are obtained by substituting $p = 0$, and $p = -\beta''$ respectively in eqn. (30). Thus at one resonant frequency near to f_0 the ring-wave ratio has its maximum value, and at the other resonant frequency this ratio is less than unity.

The analysis for $q_2 - q_1$, α'' and β'' each finite, is not given, but the two cases discussed with $q_1 = q_2$ and α'' or β'' zero independently show how the resonance Q-factors are modified.

Summarizing: inequality of propagation coefficients does not influence the variation in ring-wave ratio observed with external excitation, but the reflection and transmission in the waveguide are modified. Attenuation of the forward wave appears to counteract the effect of reflection by reducing the frequency separation of the resonances, ultimately to zero, to give a single resonance at f_0. With external excitation, the amplitude of the reverse wave exceeds that of the forward wave. Thus, for example, in the accelerator application it would appear preferable, provided low attenuation is obtained, to accelerate particles with the reverse rather than the forward wave.

Phase shift of the forward wave increases the frequency separation of the resonances without changing the Q-factors, and at resonance the ring-wave ratio is thereby modified. With sufficient phase shift the travelling-wave conditions can be approached.

(6) SOME EXPERIMENTAL RESULTS

Fig. 4 shows a waveguide ring constructed for an experimental investigation of travelling-wave resonance. A two-slot directional coupler was used, which could be changed by removing

* The contribution due to ring attenuation, i.e. q_0, and also q_c, remains unaltered.

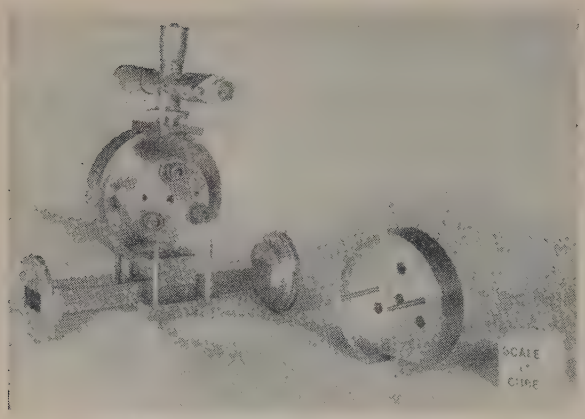


Fig. 4.—Experimental waveguide ring.

one half of the waveguide ring. The radial electric field in the ring was detected by a travelling probe located in a slot separating the two halves, and the ring-wave ratio was found from the v.s.w.r. in the normal manner. The waveguide was size No. 16, and the mean perimeter of the ring was 17·9 cm; therefore resonances were expected at guide wavelengths of 5·96 cm, 4·47 cm and 3·58 cm, corresponding to n equal to 3, 4, and 5 respectively.

With a 10 dB coupling block, a double-peaked resonance was found at 3·20 cm corresponding to $\lambda_g = 4{\cdot}48$ cm and hence $n = 4$. This was confirmed by measurement of the ring wavelength in angular degrees. The variation in ring-wave ratio and input-voltage reflection coefficient is shown in Fig. 5, plotted against the reading of a wavemeter. These curves had the form predicted by the theory, but some asymmetry was evident. The asymmetry might have been due to a variation in effective coupling through the directional coupler with position of the standing wave in the ring relative to the slots, but this feature was not investigated.

Comparing the experimental results with theory, the maximum ring-wave ratio was 3·16, and the bandwidth of the curve at 0·707 of the maximum was 0·008 72 cm, thus giving, from eqn. (31),

$$\frac{1}{2}\left(\frac{1}{Q_1}+\frac{1}{Q_2}\right) = 2\left(\frac{\lambda_g}{\lambda}\right)^2 \frac{\delta\lambda}{\lambda} = \left(\frac{4{\cdot}48}{3{\cdot}20}\right)^2 \frac{0{\cdot}008\,72}{3{\cdot}20} = \frac{1}{187}$$

It was not possible to determine any difference between Q_1 and Q_2, so that we might assume $Q_1 = Q_2 = 187$. For a coupling factor of 10 dB, the external Q-factor calculated from eqn. (27) was

$$Q_c = 8\pi/(2{\cdot}303 \log_{10} 1{\cdot}11) = 241$$

The measured loaded Q-factor and the calculated external Q-factor gave an internal Q-factor of 852, which was a low value compared with that estimated from the theoretical guide attenuation. Referring to Appendix 10.3, the relation between the frequencies at the maximum and minimum values of the input reflection coefficient given by eqn. (61) was not satisfied, the p values being 0, −1·36, and 1·09 wavemeter divisions. The check of eqn. (56) between the curve bandwidth at 0·707 of the maximum height and the bandwidth at unit wave ratio was $\sqrt{(R_m^2 - 1)} = 3$, whilst the ratio p''/p' was 3·11. These discrepancies resulted from the asymmetry of the curve, mentioned previously.

Two further experiments may be mentioned briefly. With a 20 dB coupling block, travelling-wave conditions in the ring were almost obtained, the v.s.w.r. being 0·9 at resonance; thus it appeared that, with the 10 dB coupler, the reflection causing splitting of the resonances arose at the coupler.

Some measurements were carried out with one slot of the 10 dB coupler partially blocked, to obtain semi-directivity. With

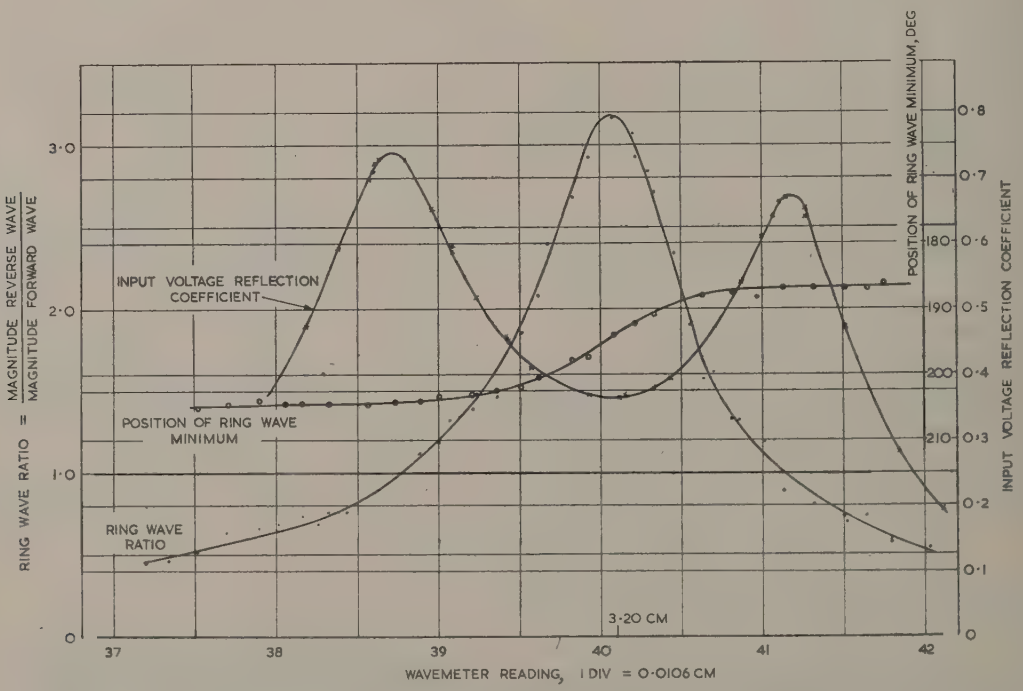


Fig. 5.—Measured ring-wave ratio and input-voltage reflection coefficient with 10 dB directional coupler.

arms 2, 3 and 4 of the coupler terminated in matched loads, the wave ratio T_3/T_4 was 0·75, and the input reflection coefficient was 0·3. These ratios varied appreciably in the frequency band of the ring resonance. Hence, a symmetrical curve of ring-wave ratio was not obtained, but its general behaviour was in accordance with that described in Appendix 10.4.

(7) CONCLUSIONS

The theory of travelling-wave resonance given in the paper shows how the wave ratio in a waveguide ring fed through a directional coupler varies in the neighbourhood of resonance.

Experimentally, the use of a moving probe in the ring has clearly illustrated the properties of travelling-wave resonance. In the case investigated, i.e. that of equal ring-wave propagation coefficients, the experimental behaviour was in accordance with the theory.

(8) ACKNOWLEDGMENTS

The work was carried out under a contract placed by the Admiralty, to whom acknowledgment is made for permission to publish the paper. The author thanks his colleagues for advice and assistance.

(9) REFERENCES

(1) TISCHER, F. J.: 'Resonance Properties of Ring Circuits', *Transactions of the Institute of Radio Engineers*, 1957, **MTT–5**, p. 51.
(2) TISCHER, F. J.: 'Resonant Properties of Nonreciprocal Ring Circuits', *ibid.*, 1958, **MTT–6**, p. 66.
(3) MILOSEVIC, L. J., and VAUTEY, R.: 'Traveling Wave Resonators', *ibid.*, 1958, **MTT–6**, p. 136.
(4) SFERRAZZA, P. J.: 'Traveling Wave Resonator', *Tele-Tech and Electronic Industries*, November, 1955, p. 84.
(5) SAXON, G.: 'Theory of Electron Beam Loading in Linear Accelerators', *Proceedings of the Physical Society*, 1954, **67 B**, p. 705.
(6) WILLSHAW, W. E., *et al.*: 'The High Power Pulsed Magnetron Development and Design for Radar Applications', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 985.
(7) PIERCE, J. R.: 'Traveling Wave Tubes' (Van Nostrand, 1950).
(8) DUNSMUIR, R.: 'The Theory of Circular Magnetrons with Uniformly Rotating Space Charge', *Proceedings I.E.E.*, Paper No. 2639 R, May, 1958 (**105 B**, Suppl. No. 10, p. 431).

(10) APPENDICES

(10.1) Derivation of Expressions for the Ring-Wave Ratio in Terms of the Resonance Q-Factors and Frequency Separation

Introducing the notation $q = n\pi/Q$, $s = 2n\pi S$, where $S = \delta f/f_0$, addition and subtraction of eqns. (29) gives

$$\left.\begin{aligned} r &= \varepsilon^E \sinh F \\ t &= \varepsilon^E \cosh F \\ \text{where} \quad E &= q_0 + q_c - \frac{q_1+q_2}{2} + j\frac{s_1+s_2}{2} \\ \text{and} \quad F &= \frac{q_2-q_1}{2} + j\frac{s_1-s_2}{2} \end{aligned}\right\} \quad . \quad . \quad (44)$$

$$\text{Therefore} \qquad t^2 - r^2 = \varepsilon^{2E}$$

$$\text{and} \qquad \frac{t}{\sqrt{(t^2-r^2)}} = \cosh F \quad . \quad . \quad . \quad . \quad (45)$$

Omitting terms of higher order than the first in the series expansions for ε^E and $\sinh F$, etc., the following approximate expressions are obtained for r and t:

$$\left.\begin{aligned} r &= \varepsilon^{q_0+q_c}\left(1 - \frac{q_1+q_2}{2} + j\frac{s_1+s_2}{2}\right)\left(\frac{q_2-q_1}{2} + j\frac{s_1-s_2}{2}\right) \\ \text{and} \quad t &= \varepsilon^{q_0+q_c}\left(1 - \frac{q_1+q_2}{2} + j\frac{s_1+s_2}{2}\right) \end{aligned}\right\} \quad (46)$$

Using the notation $p = 2n\pi P$, where $P = \Delta f/f_0$, and Δf is the frequency deviation from the mean frequency of the two resonances, we find that

$$\Gamma = q_0 + j\left(p + \frac{s_1+s_2}{2}\right) \quad . \quad . \quad . \quad . \quad (47)$$

Thus, substituting for r, t, and ε^Γ in eqn. (13), the expression obtained for the wave ratio in section 1 of the ring is

$$\left(\frac{T_5}{T_3}\right)_5 = \frac{\dfrac{q_2-q_1}{2} + j\dfrac{s_1-s_2}{2}}{\dfrac{q_1+q_2}{2} - \dfrac{1}{2}\left(\dfrac{q_2-q_1}{2}\right)^2 + j\left[p - \dfrac{1}{2}\left(\dfrac{s_1-s_2}{2}\right)^2\right]}\varepsilon^E \quad . \quad . \quad . \quad . \quad (48)$$

Eqn. (30) is obtained by neglecting the terms $\frac{1}{2}\left(\frac{q_2-q_1}{2}\right)^2$ and $\frac{1}{2}\left(\frac{s_1-s_2}{2}\right)^2$ in the above equation.

Similarly the wave ratio in section 2 of the ring is

$$\left(\frac{T_4}{T_6}\right)_6 = \frac{\left(\dfrac{q_2-q_1}{2} + j\dfrac{s_1-s_2}{2}\right)}{\dfrac{q_1+q_2}{2} + jp}\varepsilon^{-q_c} \quad . \quad . \quad (49)$$

(10.2) Power Gain in the Ring Circuit

In the testing of waveguide components and the acceleration of electrons, the power circulation which can be obtained in the ring circuit is of prime importance.

With no reflection in the ring, the gain in the wave amplitude, T_3, is

$$\frac{T_3}{I_1} = \frac{a}{1 - b\varepsilon^{-\Gamma}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

The reflection in the waveguide is zero and the transmission coefficient, τ, is

$$\tau = \frac{b\varepsilon^\Gamma - 1}{\varepsilon^\Gamma - b} \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

With low attenuation in the ring and weak coupling to the waveguide, b and Γ may be expressed in terms of the resonance Q-factors and frequency, as shown in Section 3.

Eqns. (50) and (51) then become

$$\left.\begin{aligned} \frac{T_3}{I_1} &= \frac{\sqrt{(2q_c)}}{q_0 + q_c + jp} \\ \tau &= \frac{q_0 - q_c + jp}{q_0 + q_c + jp} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (52)$$

At resonance, $p = 0$, and the power gain, proportional to $\left|\frac{T_3}{I_1}\right|^2$, is $2q_c/(q_0 + q_c)$.

The maximum possible power gain is found to be $1/2q_0$, i.e.

$Q_0/2n\pi$, which is obtained when $q_0 = q_c$. With this optimum condition, τ is zero at resonance. Otherwise transmission occurs at and off resonance. We note that low attenuation is necessary to obtain high gain, and that gain is obtained at the expense of bandwidth. Finally, we find that, with directional attenuation, the optimum coupling condition becomes $q_c = q_0 + \alpha''$.

(10.3) Derivation of the Resonance Parameters with Unequal Q-Factors

If the ring discontinuity is lossy, Q_1 and Q_2 are not equal, and to evaluate these factors it is necessary to obtain more data than are provided by the curve of ring-wave ratio since Q_1, Q_2 and S cannot all be derived from eqn. (31).

The information provided by that curve is summarized in the following equations. Using the abbreviated notation of Appendix 10.1 we have

$$q_1 + q_2 = 2p' \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

$$R_m^2(q_1 + q_2)^2 = (q_2 - q_1)^2 + s^2 \quad . \quad . \quad . \quad (54)$$

and

$$(q_1 + q_2)^2 + (2p'')^2 = (q_2 - q_1)^2 + s^2 \quad . \quad . \quad (55)$$

Subtracting eqns. (53) and (54), and substituting for $(q_2 - q_1)$ from eqn. (55), we have

$$(R_m^2 - 1)(2p')^2 = (2p'')^2 \quad . \quad . \quad . \quad . \quad (56)$$

and

$$4q_1q_2 = s^2 - (2p'')^2 \quad . \quad . \quad . \quad . \quad . \quad (57)$$

Multiplying eqn. (53) by q_1 or q_2, and substituting for q_1q_2 from eqn. (57), a solution for q_1 and q_2 may be obtained as follows:

$$q_{1,2} = \tfrac{1}{2}\left\{2p' \pm \sqrt{[(2p')^2 + (2p'')^2 - s^2]}\right\} \quad . \quad . \quad (58)$$

Thus, the Q-factors can be found when the frequency separation of the resonances is known. This separation is approximately equal to the frequency separation between the maxima in the ring-wave amplitude, or between the maxima in the input-voltage reflection coefficient, ρ, measured in the waveguide. Measurement of ρ provides the most reliable data, so we consider analysis of the ρ/frequency curve. Referring to eqn. (10) we find that

$$|\rho| \propto \left|\frac{1}{[1 - b(t + r)\varepsilon^{-\Gamma}][1 - b(t - r)\varepsilon^{-\Gamma}]}\right|$$

Substituting for b, r, t and ε^{Γ} in terms of q, s and p from Appendix 10.1, the proportionality is

$$|\rho| \propto \frac{1}{\left[q_1^2 + \left(p - \frac{s}{2}\right)^2\right]\left[q_2^2 + \left(p + \frac{s}{2}\right)^2\right]} \quad . \quad . \quad (59)$$

Differentiating eqn. (59) with respect to p and equating to zero, we have the following equation in p for the maximum and minimum values of $|\rho|$:

$$p^3 + p\left[\frac{q_2^2 + q_1^2}{2} - \left(\frac{s}{2}\right)\right] + \frac{s}{2}\left[\frac{q_1^2 - q_2^2}{2}\right] = 0 \quad . \quad (60)$$

First we note from eqn. (60) that if $q_1 = q_2$ the turning values of $|\rho|$ are $p = 0$, and $p = \pm\sqrt{[(s/2)^2 - q^2]}$.

Thus, the fractional frequency separation between the peaks is always less than S unless there is no external coupling and the internal losses are zero.

In the general case $|\rho|$ has maxima at frequencies corresponding to $p = \pm s/2$ approximately, and a minimum at approximately $p = 0$.

Denoting these frequencies by p_1, p_2 and p_3, we have, from eqn. (60),

$$p_1 + p_2 + p_3 = 0 \quad . \quad . \quad . \quad . \quad . \quad (61)$$

$$p_1p_2 + p_2p_3 + p_3p_1 = \frac{q_2^2 + q_1^2}{2} - \left(\frac{s}{2}\right)^2 \quad . \quad . \quad (62)$$

and

$$p_1p_2p_3 = \frac{s}{4}(q_2^2 - q_1^2) \quad . \quad . \quad . \quad . \quad . \quad (63)$$

From eqns. (53) and (57),

$$\tfrac{1}{2}(q_1^2 + q_2^2) = \tfrac{1}{2}[(2p')^2 + (2p'')^2 - 2(s/2)^2]$$

so that $s/2$ can be obtained from eqn. (62) as follows:

$$(s/2)^2 = \tfrac{1}{4}[(2p')^2 + \tfrac{1}{2}(2p'')^2] - \tfrac{1}{2}(p_1p_2 + p_2p_3 + p_3p_1) \quad . \quad (64)$$

q_1 and q_2 can then be found from eqn. (58). We also see that, from eqns. (58) and (63), the difference $q_1 - q_2$ is

$$q_1 - q_2 = \sqrt{[(2p')^2 + (2p'')^2 - s^2]} = -\frac{2p_1p_2p_3}{\frac{s}{2}(2p')} \quad . \quad (65)$$

The agreement between the experimental turning points of $|\rho|$ and theory can be checked by eqn. (61), and the consistency between R_m, p' and p'' can be assessed by eqn. (56).

(10.4) The Ring-Wave Ratio with Semi-Directional Coupler

The ring-wave ratio in section 1 of the ring is

$$\frac{T_5}{T_3} = r\varepsilon^{-\Gamma_1' - \Gamma_1} + \left(\frac{T_4}{T_3}\right)t\varepsilon^{-\Gamma_2} \quad . \quad . \quad . \quad (66)$$

where

$$\frac{T_4}{T_3} = \frac{br\varepsilon^{-\Gamma_1' - 2\Gamma_1} + k(1 - bt\varepsilon^{-\Gamma_2' - \Gamma_2})}{kbr\varepsilon^{-\Gamma_2' - 2\Gamma_2} + (1 - bt\varepsilon^{-\Gamma})} \quad . \quad . \quad (67)$$

with $r_1 = 0$.

The expression for T_5/T_3 simplifies if we assume that the discontinuity and the coupler divide the ring into equal sections so that $\Gamma_1 = \Gamma_2 = \Gamma/2$. With $\Gamma_1' = \Gamma_2' = 0$, the ring-wave ratio at terminal 5 is

$$\left(\frac{T_5}{T_3}\right)_5 = r + t\left[\frac{k\varepsilon^{\Gamma} - b(kt - r)}{\varepsilon^{\Gamma} - b(t - kr)}\right] \quad . \quad . \quad . \quad (68)$$

Similarly, we find that the wave ratio at terminal 6 in section 2 of the ring is

$$\left(\frac{T_4}{T_6}\right)_6 = \frac{k\varepsilon^{\Gamma} - b(kt - r)}{(t + kr)\varepsilon^{\Gamma} - b(t^2 - r^2)} \quad . \quad . \quad . \quad (69)$$

Eqns. (68) and (69) become identical with eqns. (13) and (14) respectively when $k = 0$. If we assume that the discontinuity is lossless, then from Appendix 10.5.1, eqns. (75), we find that the expressions for $kt - r$, $t + kr$, and $t^2 - r^2$ are those given in Fig. 6.

If s is small, then $|kt - r| < |t + kr|$. Therefore the position of the numbers $kt - r$, $t + kr$, and $t^2 - r^2$ in the complex plane is as shown in Fig. 6. With b real, $b(kt - r)$ and $b(t^2 - r^2)$ lie in the first quadrant and hence the magnitude of the wave ratio given by eqn. (69) is the ratio of the vectors $O'A = k\varepsilon^{\Gamma} - b(kt - r)$ and $B'C = (t + kr)\varepsilon^{\Gamma} - b(t^2 - r^2)$.

The vectors $k\varepsilon^{-\Gamma}$ and $(t + kr)\varepsilon^{-\Gamma}$ rotate in a clockwise

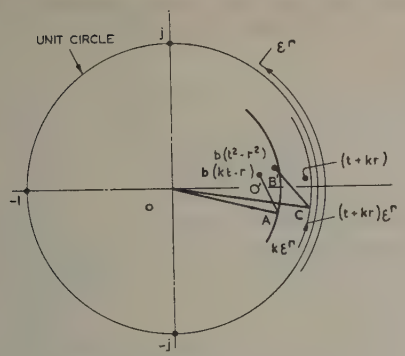


Fig. 6.—Vector diagram for ring-wave ratio with semi-directional coupler, and for lossless discontinuity.

$$\text{Ring-wave ratio} = \frac{\text{O'A}}{\text{B'C}} = \frac{|k\varepsilon^{\Gamma} - b(kt-r)|}{|(t+kr)\varepsilon^{\Gamma} - b(t^2-r^2)|}$$

For lossless discontinuity,

$$kt - r = \sqrt{(k^2\cos^2\tfrac{1}{2}s + \sin^2\tfrac{1}{2}s)}\,\varepsilon^{j[\frac{1}{2}s + \arctan(\tan\frac{1}{2}s/k)]}$$

$$t + kr = \sqrt{(\cos^2\tfrac{1}{2}s + k^2\sin^2\tfrac{1}{2}s)}\,\varepsilon^{j[\frac{1}{2}s + \arctan(k\tan\frac{1}{2}s)]}$$

$$t^2 - r^2 = \varepsilon^{js}$$

direction with increasing frequency, and the diagram shows that the wave ratio has a single maximum in the neighbourhood of $\mathscr{I}\Gamma = 0$.

(10.5) Voltage Patterns of the Ring Resonances

First we consider the voltage patterns of resonances in a closed ring. Inserting the appropriate boundary conditions in eqn. (2), we obtain three expressions for the wave ratio at the discontinuity as follows:

$$\frac{T_5}{T_6} = \frac{r\varepsilon^{-\Gamma'}}{\varepsilon^{\Gamma} - t} = \frac{\varepsilon^{\Gamma} - t\varepsilon^{-\Gamma'}}{r} = \pm\sqrt{\left(\varepsilon^{\Gamma'}\frac{1 - t\varepsilon^{-\Gamma'-\Gamma}}{1 - t\varepsilon^{-\Gamma}}\right)} \quad . \quad (70)$$

Eqn. (70) shows that all the resonances have a standing-wave field pattern if Γ' is zero, with either a voltage maximum or a voltage zero at the discontinuity, corresponding to $T_5/T_6 = 1$ or -1, respectively.

If we represent the discontinuity by a shunt impedance, the frequency and Q-factors of the resonance with $T_5/T_6 = -1$ are those of the plain ring, since no current flows into the impedance.

The converse is true for the resonance with a voltage maximum at the impedance, and hence the frequency or the Q-factor, or both, may be different from that of the plain ring.

With $\Gamma' = 0$ we have shown that the equation giving the resonant frequencies and Q-factors is, in this case, $\varepsilon^{\Gamma_{1,2}} = t \pm r$.

Referring to Appendix 10.5.1, we see that, for a shunt admittance, $t - r = 1$ and $t + r = 1 + 2r$. Thus, for one resonance, corresponding to the solution $\varepsilon^{\Gamma} = t - r = 1$, the frequency and Q-factor are those of the plain ring. For the other resonance, corresponding to $\varepsilon^{\Gamma} = 1 + 2r$, both frequency and Q-factor may be different from f_0 and Q_0 respectively, depending upon whether the impedance is reactive, resistive or both.

For a shunt resistance, $\arg(t + r)$ is zero, so that both resonances have the same frequency, f_0, but different Q-factors. For a shunt reactance, $|t + r|$ is unity so that both resonances have the same Q-factor, Q_0, but the frequency of one resonance, corresponding to $\arg(t + r) \neq 0$, is different from f_0. For a shunt impedance, $|t + r| \neq 1$ and $\arg(t + r) \neq 0$, so that both the frequency and Q-factor of one resonance differ from f_0 and Q_0 respectively.

For a series impedance, $(t + r) = 1$, and $(t - r) = 1 - 2r$. The relation between frequency and Q-factor may be found for series resistance, reactance or impedance in the way described for the corresponding shunt values, and the results are identical. However, the field pattern of the resonances is displaced by $\lambda_g/4$ relative to the pattern with a shunt impedance, so that maximum current and zero voltage occur at the series impedance for the resonance with modified frequency and Q-factor.

Summarizing: the wave pattern at the resonant frequencies is oriented so that there is coupling or no coupling with the discontinuity, depending upon whether the waves are in phase or in phase opposition at the discontinuity.

With unequal ring-wave propagation coefficients, and taking account of the directional coupler, the ring-wave ratio at resonance is found to be

$$\left(\frac{T_5}{T_3}\right)_5 = \left[\frac{\dfrac{q_2 - q_1}{2} + j\dfrac{s}{2}}{-\dfrac{\Gamma'}{2} \pm \phi}\right]\varepsilon^{E} \quad . \quad . \quad . \quad (71)$$

Detailed investigation of this equation shows that both attenuation and phase shift of the forward wave result in unequal wave magnitudes, and reorientation of the field pattern. The change in the resonance frequencies and Q-factors is discussed in Section 5.1.

(10.5.1) Expressions for r and t for Shunt and Series Impedances in Terms of the Resonance Q-factors and Frequency Separation.

For a normalized admittance $y = 1/z$ shunted across a transmission line, the voltage reflection and transmission coefficients are $r = -y/(2 + y)$, and $t = 2/(2 + y)$.

$$\left.\begin{aligned} &\text{Therefore} && t - r = 1 \\ &&& t + r = 1 + 2r \\ &\text{and} && t \pm r = \frac{2 \mp y}{2 + y} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (72)$$

Putting $b = 1$, the ring resonance equation (21) can be written $\varepsilon^{-q_x + js} = t \pm r$. The contribution to the loading of the resonances, and their frequency separation, is found from

$$\varepsilon^{-q_x + js} = \sqrt{\left[\frac{(2-g)^2 + b^2}{(2+g)^2 + b^2}\right]}\varepsilon^{-\arctan\left[4b/(4 - g^2 - b^2)\right]} \quad . \quad (73)$$

For a shunt conductance, g, the solution of eqn. (73) is $g = 2\tanh q_x/2$ and hence

$$\left.\begin{aligned} &r = -\tfrac{1}{2}(1 - \varepsilon^{-q_x}) \\ &t = \tfrac{1}{2}(1 + \varepsilon^{-q_x}) \\ &\text{with} \quad \sqrt{(t^2 - r^2)} = \varepsilon^{-q_x/2} \\ &\text{and} \quad \frac{t}{\sqrt{(t^2 - r)}} = \cosh\frac{q_x}{2} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (74)$$

Similarly, for a shunt susceptance b, the solution of eqn. (73) for s is $s = -2\arctan b/2$ and therefore

$$\left.\begin{aligned} r &= -j\sin\frac{s}{2}\varepsilon^{j\frac{s}{2}} \\ t &= \cos\frac{s}{2}\varepsilon^{j\frac{s}{2}} \\ \sqrt{(t^2 - r^2)} &= \varepsilon^{j\frac{s}{2}} \\ \frac{t}{\sqrt{(t^2 - r^2)}} &= \cos\frac{s}{2} \end{aligned}\right\} \quad . \quad . \quad . \quad (75)$$

For a shunt admittance, $y = g + jb$, we find that approximate solutions of eqn. (73) are $b = -s$ and $g = q_x$ when s and q_x are small. Therefore the reflection and transmission coefficients are

$$\left.\begin{aligned} r &= \frac{-q_x + js}{2 + q_x - js} \simeq \frac{-q_x + js}{2} \\ t &= \frac{2}{2 + q_x - js} \simeq 1 - \frac{q_x}{2} + j\frac{s}{2} \end{aligned}\right\} \quad . \quad . \quad (76)$$

Putting $q_0 = q_c = q_2 = 0$, and $q_1 = q_x$ in the equations for r and t given in Appendix 10.1, we find those expressions become identical with eqn. (76). For a series impedance, z, the voltage reflection and transmission coefficients are $r = z/(1 + z)$ and $t = 2/(2 + z)$.

Therefore
$$\left.\begin{aligned} t + r &= 1 \\ t - r &= 1 - 2r \end{aligned}\right\}$$
and
$$\left.\begin{aligned} t \pm r &= \frac{2 \pm z}{2 + z} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (77$$

We find that the equations for r and t are unchanged except that for a series resistance, $r = \frac{1}{2}(1 - \varepsilon^{-q_x})$, and for a series reactance $t = j \sin \frac{1}{2}s\varepsilon^{j\frac{1}{2}s}$.

621.375.9 : 621.372.2

The Institution of Electrical Engineers
Paper No. 3017 E
Oct. 1959
©

SATURATION EFFECTS IN A TRAVELLING-WAVE PARAMETRIC AMPLIFIER

By A. JURKUS, B.Sc.A., and P. N. ROBSON, B.A., Associate Member.

(*The paper was first received 4th February, and in revised form 16th April, 1959. It was published in October, 1959, and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION *9th November, 1959.*)

SUMMARY

The equations governing the voltage distribution along a parametric travelling-wave amplifier can, under certain conditions, be solved in terms of elliptic functions.

The solution shows that there is a spatially periodic transfer of energy between the signal, idling and pumping waves. The maximum saturated power gain that can be achieved is

$$1 + \frac{\omega_1 \text{ pump input power}}{\omega_3 \text{ signal input power}}$$

where ω_1 and ω_3 are the angular frequencies of the signal and of the pump, respectively. The positions of these gain maxima can be calculated explicitly without invoking small-signal conditions.

LIST OF SYMBOLS

C = Capacitance per unit length.
C_0 = Small-signal value of C.
K = Complete elliptic integral of the first kind.
k = Modulus of elliptic function.
k' = Complementary modulus of elliptic function.
l_{max} = Distance of first saturated gain maximum.
P_1, P_2, P_3 = Power in signal, idling and pumping waves.
V_1, V_2, V_3 = Amplitude of signal, idling and pumping waves.
x = Distance along transmission line.
α_0, α_1 = Constants associated with the growth of the amplitudes.
$\beta_1, \beta_2, \beta_3$ = Phase constants of signal, idling and pumping waves.
η = Constant of proportionality between capacitance change and applied voltage.
$\lambda_1, \lambda_2, \lambda_3, \lambda_p$ = Wavelengths of voltage and power transfer between the three waves.
$\omega_1, \omega_2, \omega_3$ = Angular frequencies of signal, idling and pumping waves.

(1) INTRODUCTION

A travelling-wave parametric amplifier in which interaction between pump, idling and signal voltages is achieved along a transmission line with distributed non-linear capacitance has been described by Cullen.[1] The voltage equations [eqns. (13)–(18)] of Cullen's paper are here taken as the starting point. The usual assumption is made that the second derivative of the amplitude of the voltage wave with respect to distance can be neglected. The resulting equations can then be solved exactly and the solutions are discussed in some detail. Cullen gives only approximate solutions of these equations.

(2) BRIEF SUMMARY OF BASIC THEORY

The model is assumed to be a lossless transmission line having constant distributed inductance, L_0, per unit length and a distributed non-linear capacitance per unit length which obeys the simplified law, $C = C_0(1 + \eta v)$. It is further assumed that only three frequencies need be considered. These are the signal frequency, ω_1, idling frequency, ω_2, and pump frequency, ω_3, related by

$$\omega_1 + \omega_2 = \omega_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It is shown in the Appendix that, if the upper idling frequency, $\omega_4 = \omega_3 + \omega_1$, is also included, there is no gain at the signal frequency.

Following Cullen,[1] the voltage wave is written

$$v = V_1 \cos(\omega_1 t - \beta_1 x + \phi_1) + V_2 \cos(\omega_2 t - \beta_2 x + \phi_2) + V_3 \cos(\omega_3 t - \beta_3 x + \phi_3)$$

where V_1, V_2 and V_3 are functions of x, and the phase coefficients β_1, β_2 and β_3 are assumed to be related by

$$\beta_1 + \beta_2 = \beta_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Cullen then derives the following equations to be satisfied:

$$\frac{d^2V_1}{dx^2} = -\eta\beta_1^2 V_2 V_3 \cos\psi \quad . \quad . \quad . \quad . \quad (3)$$

$$\frac{dV_1}{dx} = -\frac{\eta}{2}\beta_1 V_2 V_3 \sin\psi \quad . \quad . \quad . \quad . \quad (4)$$

$$\frac{d^2V_2}{dx^2} = -\eta\beta_2^2 V_3 V_1 \cos\psi \quad . \quad . \quad . \quad . \quad (5)$$

$$\frac{dV_2}{dx} = -\frac{\eta}{2}\beta_2 V_3 V_1 \sin\psi \quad . \quad . \quad . \quad . \quad (6)$$

$$\frac{d^2V_3}{dx^2} = -\eta\beta_3^2 V_1 V_2 \cos\psi \quad . \quad . \quad . \quad . \quad (7)$$

$$\frac{dV_3}{dx} = +\frac{\eta}{2}\beta_3 V_1 V_2 \sin\psi \quad . \quad . \quad . \quad . \quad (8)$$

where $\psi = \phi_1 + \phi_2 - \phi_3$.

(3) SOLUTION OF EQUATIONS

If we consider eqns. (3)–(8), which are not self-consistent, we notice that self-consistency is ensured by assigning to ψ a value such that $\cos\psi = 0$. Alternatively we can consider that this amounts to neglecting second derivatives, i.e. considering as solutions only slowly varying functions. Since the resulting solutions are shown to be in agreement both with the Manley–Rowe relations and with power conservation, the procedure seems to be justifiable.

We put thus $\sin\psi = \pm 1$, and consider only the positive sign; it can be clearly seen from the solutions that taking the negative sign is equivalent to reversing the phase of the idling-wave amplitude. The equations to be solved then reduce to:

$$\frac{dV_1}{dx} = -\frac{\eta}{2}\beta_1 V_2 V_3 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\frac{dV_2}{dx} = -\frac{\eta}{2}\beta_2 V_3 V_1 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\frac{dV_3}{dx} = +\frac{\eta}{2}\beta_3 V_1 V_2 \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Mr. Jurkus and Mr. Robson are in the Department of Electrical Engineering, University of Sheffield.

These equations have the same form as Euler's equations of motion for a rigid body moving about a fixed point with no external forces applied. In this case, time t is replaced by distance x, and the three components of angular velocity about the principal axes by the amplitudes of the pump, idling and signal voltages V_3, V_2 and V_1, respectively.

From eqns. (9) and (10),

$$\frac{dV_1}{dV_2} = \frac{\beta_1 V_2}{\beta_2 V_1}$$

Integrating
$$V_1^2 = \frac{\beta_1}{\beta_2} V_2^2 + \text{constant} \quad . \quad . \quad . \quad . \quad (12)$$

When the idling voltage V_2 is zero at $x = 0$, the signal voltage has its initial value V_{10}. Hence eqn. (12) may be written

$$V_1^2 = \frac{\beta_1}{\beta_2} V_2^2 + V_{10}^2 \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Similarly, taking eqns. (10) and (11) together, and eqns. (9) and (11) together,

$$V_3^2 = -\frac{\beta_3}{\beta_2} V_2^2 + V_{30}^2 \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and
$$V_1^2 = -\frac{\beta_1}{\beta_3}(V_3^2 - V_{30}^2) + V_{10}^2 \quad . \quad . \quad . \quad (15)$$

Substituting eqns. (13) and (14) into eqn. (10) and taking only the positive square roots in both cases,

$$\frac{dV_2}{dx} = -\frac{\eta}{2}\left[\beta_1\beta_3\left(\frac{\beta_2}{\beta_3}V_{30}^2 - V_2^2\right)\left(\frac{\beta_2}{\beta_1}V_{10}^2 + V_2^2\right)\right]^{1/2} \quad . \quad (16)$$

Writing
$$A^2 = \frac{\beta_2}{\beta_3} V_{30}^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$B^2 = \frac{\beta_2}{\beta_1} V_{10}^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

eqn. (16) becomes, putting in the limits of integration,

$$\int_0^{V_2} \frac{dV_2}{(A^2 - V_2^2)^{1/2}(B^2 + V_2^2)^{1/2}} = -\frac{\eta}{2}\sqrt{(\beta_1\beta_3)}\int_0^x dx \quad . \quad (19)$$

The integral on the left-hand side of eqn. (19) can be reduced to the standard form of an elliptic integral of the first kind by writing $A^2 - V_2^2 = A^2T^2$.

Eqn. (19) then simplifies to

$$\frac{1}{(B^2 + A^2)^{1/2}}\int_1^{(1 - V_2^2/A^2)^{1/2}} \frac{dT}{(1 - T^2)^{1/2}(1 - k^2T^2)^{1/2}} = \frac{\eta}{2}\sqrt{(\beta_1\beta_3)}x \quad . \quad (20)$$

where
$$k^2 = \frac{A^2}{A^2 + B^2} \quad (21)$$

The integral can be written as

$$\int_1^{(1 - V_2^2/A^2)^{1/2}} = \int_0^{(1 - V_2^2/A^2)^{1/2}} - \int_0^1$$

The first integral is an incomplete elliptic integral of the first kind whose solution is $\mathrm{sn}^{-1}(1 - V_2^2/A^2)^{1/2}$ and the second is a complete elliptic integral

$$K = \mathrm{F}\left(k, \frac{\pi}{2}\right) \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Evaluating the integral in eqn. (20) and also using eqns. (17), (18), (21) and (22) we obtain, after some reduction,

$$V_2 = \sqrt{\left(\frac{\beta_2}{\beta_3}\right)} V_{30}\,\mathrm{cn}\left(K + \frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad (23)$$

where
$$\alpha_0 = \frac{\eta V_{30}}{2}\sqrt{(\beta_1\beta_2)} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

and
$$k = \left(1 + \frac{\beta_3 V_{10}^2}{\beta_1 V_{30}^2}\right)^{-1/2} \quad . \quad . \quad . \quad . \quad (25)$$

This can be reduced to one of the alternative forms

$$V_2 = -\sqrt{\left(\frac{\beta_2}{\beta_1}\right)} V_{10} k\,\mathrm{sd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad (26)$$

or
$$V_2 = -\sqrt{\left(\frac{\beta_2}{\beta_3}\right)} V_{30} k'\,\mathrm{sd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad (27)$$

where $k' = \left(1 + \frac{\beta_1 V_{30}^2}{\beta_3 V_{10}^2}\right)^{-1/2}$ is the complementary modulus of the elliptic function.

Expressions for V_1 and V_3 can now be obtained by substituting eqn. (23) in eqns. (13) and (14), whence

$$V_1 = \sqrt{\left(\frac{\beta_1}{\beta_3}\right)} V_{30}\frac{1}{k}\,\mathrm{dn}\left(K + \frac{\alpha_0}{k}x\right) \quad . \quad . \quad (28)$$

which can be reduced to

$$V_1 = V_{10}\,\mathrm{nd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

or
$$V_1 = \sqrt{\left(\frac{\beta_1}{\beta_3}\right)} \frac{k'}{k} V_{30}\,\mathrm{nd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad (30)$$

and
$$V_3 = V_{30}\,\mathrm{sn}\left(K + \frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad (31)$$

which can be reduced to

$$V_3 = V_{30}\,\mathrm{cd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

or
$$V_3 = \sqrt{\left(\frac{\beta_3}{\beta_1}\right)} \frac{k}{k'} V_{10}\,\mathrm{cd}\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad (33)$$

The form of these expressions is shown in Fig. 1. It will be seen that there is a periodic variation in the amplitudes of all three waves. If we write λ_1, λ_2, λ_3 for the spatial wavelengths of the signal, idling and pumping amplitudes, respectively, then, from eqns. (26), (29) and (32),

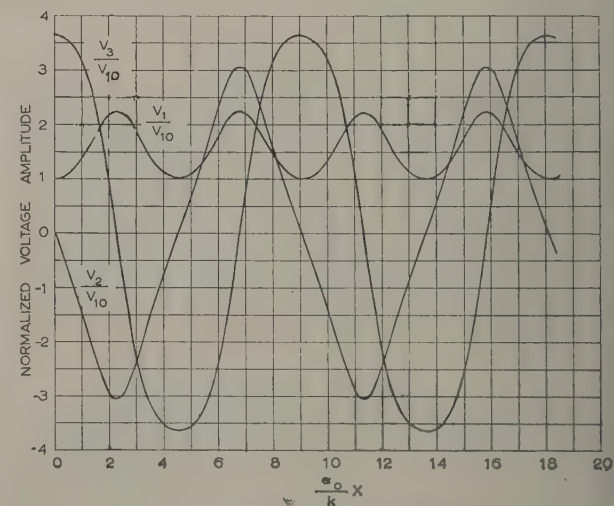


Fig. 1.—Variation of the amplitude of signal, idling and pump voltages along the line for $k^2 = 0{\cdot}8$.

$$\lambda_1 = \frac{2Kk}{\alpha_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

$$\lambda_2 = \lambda_3 = \frac{4Kk}{\alpha_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

It is interesting to note that the wavelength of the signal voltage is one-half that of the pump or idling voltage.

The power expressions can readily be obtained from eqns. (26), (29) and (32) on the assumption that the characteristic impedance of the line is the same for all three frequencies. Hence

$$P_1 = P_{10}\,\mathrm{nd}^2\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

$$P_2 = \frac{\omega_2}{\omega_1}P_{10}k^2\,\mathrm{sd}^2\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad . \quad (37)$$

$$P_3 = P_{30}\,\mathrm{cd}^2\left(\frac{\alpha_0}{k}x\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

These expressions are shown plotted in Fig. 2. The wavelength of the spatial power distributions in all three waves is now the same and equal to, say, λ_p.

i.e. $$\lambda_p = \frac{2Kk}{\alpha_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

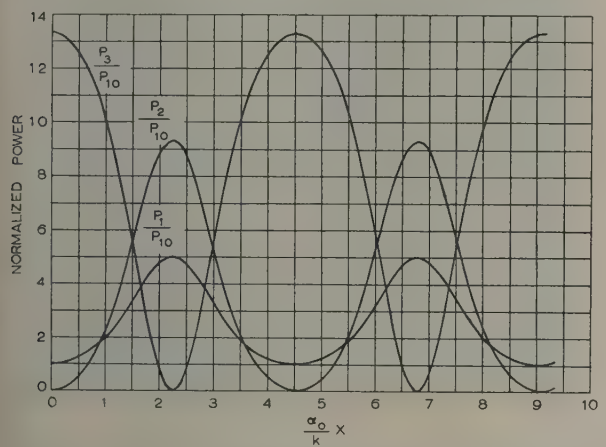


Fig. 2.—Variation of signal, idling and pump power along the line for $k^2 = 0{\cdot}8$.

If $\frac{\omega_3}{\omega_1}\frac{P_{10}}{P_{30}} \ll 1$, k tends to unity and the elliptic functions reduce to hyperbolic functions. Eqns. (36), (37) and (38) can be written under this condition as

$$P_1 = P_{10}\cosh^2(\alpha_0 x) \quad . \quad . \quad . \quad . \quad (36a)$$

$$P_2 = \frac{\omega_2}{\omega_1}P_{10}\sinh^2(\alpha_0 x) \quad . \quad . \quad . \quad (36b)$$

$$P_3 = P_{30} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36c)$$

They may be recognized as the solutions Cullen gives under small-signal conditions.

(4) CONSERVATION OF POWER AND MANLEY–ROWE RELATIONS

Since the line has been assumed lossless, the total power must be conserved; this can readily be proved from eqns. (36), (37) and (38):

$$P_1 + P_2 + P_3 = P_{10} + P_{30}$$

for all values of x.

Differentiating eqns. (36), (37) and (38) with respect to distance x, we find

$$\frac{1}{\omega_1}\frac{dP_1}{dx} = \frac{1}{\omega_2}\frac{dP_2}{dx} = -\frac{1}{\omega_3}\frac{dP_3}{dx}$$

This shows that energy is transferred between the three frequencies according to the Manley–Rowe relations.[2]

(5) SATURATION EFFECTS

It is clear from Fig. 2 that there is a maximum signal power, P_{1max}, that can be obtained, and this occurs when all the power has been extracted from the pumping wave. The power extracted from the pump is divided between the signal and idling waves through the Manley–Rowe relations. The maximum obtainable power gain can be readily obtained from eqn. (36) and is given by

$$\frac{P_{1max}}{P_{10}} = \left(1 + \frac{\omega_1 P_{30}}{\omega_3 P_{10}}\right) = \frac{1}{k'^2} \quad . \quad . \quad . \quad (40)$$

Again from eqn. (36) it can be found that the first maximum of gain occurs at a distance from the origin given by

$$l_{max} = \frac{Kk}{\alpha_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

It is instructive to observe how the position of this first maximum varies as (*a*) the gain is decreased by increasing the signal input power and keeping the pump input constant, and (*b*) the gain is increased by increasing the pump input power and keeping the signal input constant.

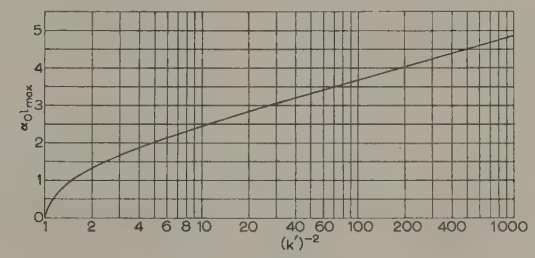


Fig. 3.—Position of first saturated gain maximum for constant pump input.

In the first case, we can use eqn. (41) directly. Fig. 3 shows $\alpha_0 l_{max}$ plotted as a function of the saturated power gain $(k')^{-2}$. It is seen that under these conditions the distance of the gain maximum from the input increases as the gain increases.

In the second case, eqn. (41) is most suitably written in the form:

$$\alpha_1 l_{max} = Kk' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where $$\alpha_1 = \frac{\eta}{2}V_{10}\sqrt{(\beta_2\beta_3)} \quad . \quad . \quad . \quad . \quad (43)$$

In Fig. 4 $\alpha_1 l_{max}$ is plotted in terms of the saturated power gain. It is seen that the distance of the first gain maximum starts at

$$l_{max} = \frac{\pi}{\eta V_{10}\sqrt{(\beta_2\beta_3)}} \quad . \quad . \quad . \quad . \quad (44)$$

when the gain is unity, and decreases to zero as the gain increases. The corresponding approximate curve given by Cullen is also plotted in this Figure. It is seen that there is a considerable difference between the two curves.

In Fig. 5 the variation of the signal power with distance is

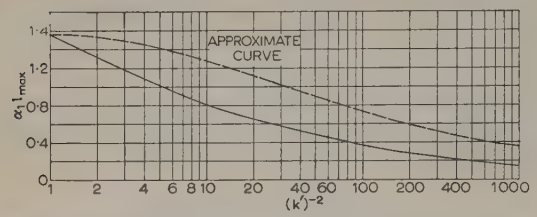


Fig. 4.—Position of first saturated gain maximum for constant signal input.

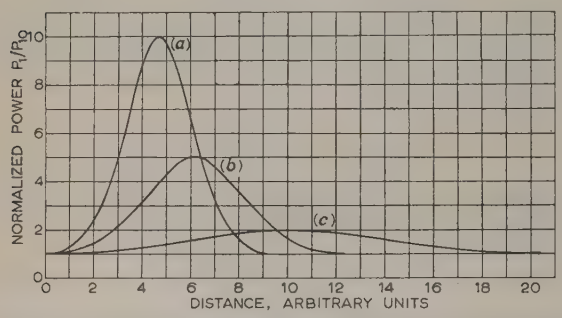


Fig. 5.—Variation of signal power along the line for three values of saturated gain.
(*a*) $k^2 = 0{\cdot}9$.
(*b*) $k^2 = 0{\cdot}8$.
(*c*) $k^2 = 0{\cdot}5$.

shown for three values of maximum gain. It can be clearly seen from this diagram that, if the amplifier were working under saturation conditions, an increase in pumping power, for example, might not necessarily result in an increase in gain, as would be the case under small-signal conditions.

Fig. 5 could also be interpreted as resulting from a change in signal frequency. It follows then that the gain of an amplifier of a given length may, under large-signal conditions, fall to unity several times within the frequency range of ω_1 from 0 to ω_3. This is to be contrasted with the gain under small-signal conditions when the gain rises from unity to a maximum at $\omega_1 = \omega_3/2$* and falls again to unity when $\omega_1 = \omega_3$.

(6) CONCLUSIONS

In the type of travelling-wave parametric amplifier considered, it has been shown that, under large-signal conditions, there is a periodic transfer of power between the signal, idling and pumping waves, respectively.

The initial eqns. (9), (10) and (11) governing the spatial variations of the three voltages have been solved exactly. However, in order to achieve this, a number of restrictions had to be imposed. The law of capacitance variation, for example, is somewhat simplified, and no account is taken of the power that will reside in the harmonics. This last restriction can, however, be justified by the fact that the practical form of the amplifier at the moment has lumped non-linear capacitances instead of the continuously distributed medium assumed in the analysis. It can be defined essentially as a low-pass filter, rejecting most of the frequencies higher than the pump. In fact, interaction between the upper sideband and the other frequencies can be prevented by other means. For instance, the dispersion characteristic can be such that the upper sideband gets rapidly out of step with the other waves.

* This point is a critical one, since conditions of synchronism must now be considered. This case is not studied in detail here.

The bandwidth of the amplifier will be limited by the dispersion characteristics of the line in small-signal operation. Under large-signal conditions an additional limitation will be imposed owing to the variation of the period of power transfer with frequency.

Measurements on the type of line described are currently being undertaken and it is hoped to publish the results later.

(7) ACKNOWLEDGMENTS

The authors are grateful to Professor A. L. Cullen for his encouragement and many helpful discussions. The work described forms part of a programme of research in microwave electronics, and the support by C.V.D. of this programme is gratefully acknowledged. Mr. Jurkus acknowledges the financial help of an Athlone Fellowship which made possible his stay in this country.

(8) REFERENCES

(1) CULLEN, A. L.: 'Theory of the Travelling-Wave Parametric Amplifier', *Proceedings I.E.E.*, Paper No. 2991 E, August, 1959 (see page 101).

(2) MANLEY, J. M., and ROWE, H. E.: 'Some General Properties of Nonlinear Elements, Part I: General Energy Relations, *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 904.

(9) APPENDIX

If both upper and lower idling frequencies are considered, the equations can be written

$$\frac{dV_1}{dx} + \frac{\eta}{2}\beta_1 V_3(V_2 + V_4) = 0 \quad . \quad . \quad . \quad (45)$$

$$\frac{dV_2}{dx} + \frac{\eta}{2}\beta_2 V_1 V_3 = 0 \quad . \quad . \quad . \quad . \quad . \quad (46)$$

$$\frac{dV_3}{dx} - \frac{\eta}{2}\beta_3 V_1(V_2 - V_4) = 0 \quad . \quad . \quad (47)$$

$$\frac{dV_4}{dx} - \frac{\eta}{2}\beta_4 V_1 V_3 = 0 \quad . \quad . \quad . \quad . \quad . \quad (48)$$

where

$$\beta_2 = \beta_3 - \beta_1$$
$$\beta_4 = \beta_3 + \beta_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

Eqns. (45)–(48) can be shown to satisfy the Manley–Rowe relations, which are now written:

$$\frac{1}{\omega_2}\frac{dP_2}{dx} + \frac{1}{\omega_3}\frac{dP_3}{dx} + \frac{1}{\omega_4}\frac{dP_4}{dx} = 0 \quad . \quad . \quad (50)$$

$$-\frac{1}{\omega_2}\frac{dP_2}{dx} + \frac{1}{\omega_1}\frac{dP_1}{dx} + \frac{1}{\omega_4}\frac{dP_4}{dx} = 0 \quad . \quad . \quad (51)$$

The solutions of eqns. (45)–(48) can be obtained in the same manner as in the three-frequency case; we shall only show here that no amplification is possible when the sum idling frequency is present.

From eqns. (46) and (48),

$$\frac{1}{\omega_2^2}\frac{dP_2}{dx} = \frac{1}{\omega_4^2}\frac{dP_4}{dx} \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Substituting this in eqn. (51),

$$\frac{1}{\omega_1}\frac{dP_1}{dx} = -\frac{2\omega_1}{\omega_2^2}\frac{dP_2}{dx}$$

Since $P_2 = 0$ at $x = 0$, dP_2/dx can only be positive. Therefore dP_1/dx is negative, and power is extracted from the signal frequency.

DISCUSSION ON THE ABOVE THREE PAPERS AND ON 'THE GAIN OF TRAVELLING-WAVE FERRO MAGNETIC AMPLIFIERS'* BEFORE THE ELECTRONICS AND COMMUNICATIONS SECTION, 9TH NOVEMBER, 1959

Mr. A. F. H. Thomson: Dr. Clarricoats showed that large amounts of pump power were required in the type of amplifier which he was considering, and the main reason for this is that the pump power is being used to excite the precession of the magnetization vector. This, however, is not the only form of amplifier which can be set up. By putting the pump field in the direction of the d.c. field and the signal and idler fields in the plane at right angles to it, parametric amplification should, ideally, be possible without exciting precession. In experiments of this sort we have, however, found a new type of absorption in these materials. With a d.c. field corresponding to half the pump frequency it was hoped to observe amplification in the non-degenerate mode with idler and signal frequencies close together. As the pump field was increased an absorption of pump power was observed above a critical pump field of about 0·5 oersted for an yttrium-iron garnet single crystal.

In Fig. A the absorption of the pump power is shown vertically

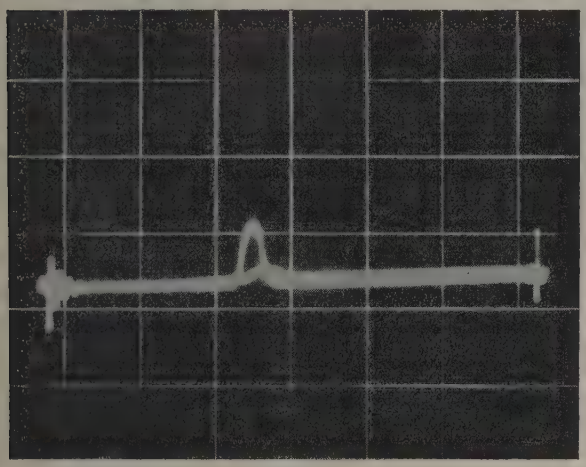


Fig. A

and a small sweep of the d.c. field horizontally, representing a sweep of some 40 oersteds over the mean of about 4000 oersteds. At almost half the pump frequency we get an absorption which appears only above a certain amplitude of the pump field. This photograph was taken with the amount of power fed in about 1% above the threshold value.

In Fig. B we have gone up to 2% above threshold power, and the absorption has increased and widened. The absorptions on the forward and return traces are not the same and for some time I thought this was due to imperfections in the experiment, but subsequently found that the form of the absorption curve depends on the rate at which the field is swept, and also on whether it is increasing or decreasing. I have not been able to determine whether the higher peak is due to a rising or falling field, but I think the former.

Fig. C shows the relationship between the absorption on an arbitrary scale and the amplitude of the r.f. field. The curve is much steeper than it looks, because the absorption has reduced the r.f. field and has almost certainly maintained it at a constant value. The ordinates are the fields which would exist had there not been any absorption. This mechanism of absorption

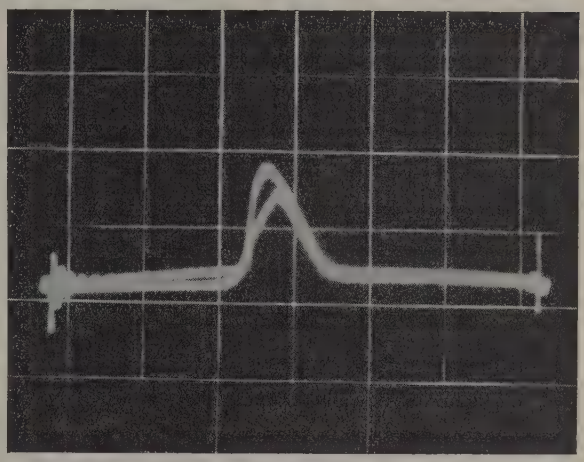


Fig. B

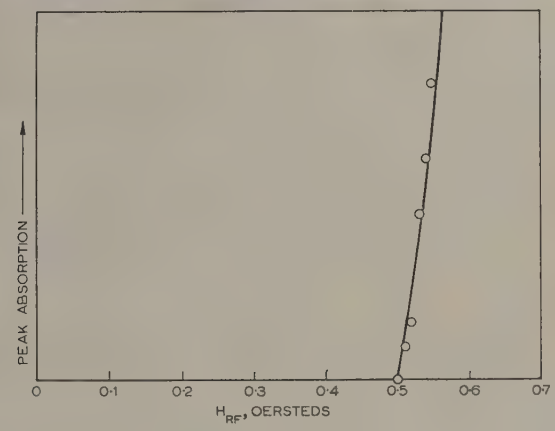


Fig. C

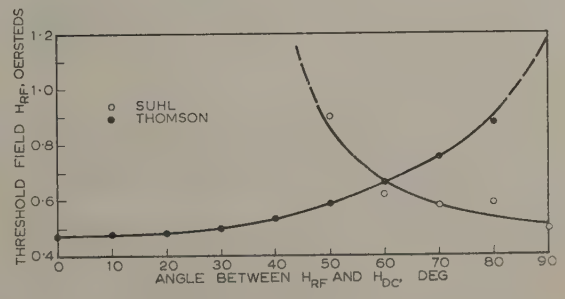


Fig. D

differs from that reported by Suhl and Walker, as is seen in Fig. D, although it is certainly due to spin-wave excitation.

The curve in Fig. D was obtained by continuously varying the angle between the pump and d.c. fields. The Suhl and Walker phenomenon, which is more difficult to observe, seems to have a minimum threshold of about the same value as ours. It differs, however, in that this subsidiary absorption occurs, not at

* CLARRICOATS, P. J. B., Monograph No. 334 E, May, 1959 (see 106 C, p. 165).

half the field appropriate to the pump frequency, but near 70%. Our curve corresponds almost exactly to a $1/\cos^2\theta$ form, which one would expect if it were merely due to the resolved component in the d.c. field direction. We think we have an explanation for the mode of excitation of these spin-wave modes, but we have not yet managed to prove it.

The time-constant of build-up of this absorption can be extremely long. It is measured in milliseconds for a power of about 2% above threshold. We now have a better method of measuring the threshold field without varying the d.c. field at all, so that a true c.w. threshold is measured. To find the order of the spin-wave modes which are excited, we have tried grinding the samples to see whether we could reach a size where the threshold field begins to increase. This would give an idea of the spin-wave-mode wavelength. We have found that this begins to go up rather suddenly at about 200 microns. This is hard to understand, because with such wave numbers we should have been able to detect radiation from the signal and idler. We have looked for it with a sensitive receiver but have never observed any radiation. There are, therefore, still a number of points to clear up. However, spin waves again limit the value of permissible pump field, as in the configuration discussed by Dr. Clarricoats.

Mr. R. W. White: With regard to Prof. Cullen's paper, I am a little disappointed that it is concerned with uniformly distributed voltage-dependent capacitance and suggest that it is of much greater importance at the moment to consider a line structure having discrete voltage-dependent shunt capacitances distributed regularly, or perhaps in some more complex pattern, along the line. Can Prof. Cullen's work be extended rigorously to such cases?

The theory of periodically-loaded parametric amplifiers deals, in general, with cases where the diodes are all identical; yet there is very great difficulty in realizing matched sets of more than three or four diodes, even when individual bias adjustments can be used. This practical restriction is particularly serious when uniformity of gain in the pass-band and good matching are required.

In theoretical treatments some simplification of the diode characteristic is undoubtedly necessary; but some variable-capacitance-diode simplifications seem a little too far away from practical realities.

Simple parametric amplifiers using a single diode have marked limitations in bandwidth and gain stability. Travelling-wave types offer much greater bandwidth, but there are still serious difficulties in their practical realization outside the laboratory. There is much to be said for some form of amplifier using two (or at most four) diodes and offering characteristics intermediate between these two types.

Finally, have any of the authors yet obtained consistent and repeatable experimental results which show conclusively that travelling-wave parametric amplifiers have really marked advantages in gain, stability of gain, bandwidth and noise factor over the best travelling-wave valves?

Mr. R. A. Waldron: There is a fact concerning ferromagnetic resonance which may not be generally appreciated, and which has a bearing on Dr. Clarricoats's amplifier. Polder has shown that for a certain value of σ, the internal polarizing field, the elements of the permeability tensor of a ferrite become infinite in the lossless case. Kittel has shown that for a certain value of σ dependent on the shape of a ferrite specimen, as well as on frequency, the absorption of energy in a cavity containing the specimen is a maximum. At the Polder value of σ, microwave energy is almost entirely excluded from the specimen, while at the Kittel value, which for most shapes of specimen is different from the Polder value, energy is concentrated in the ferrite.

The non-linear behaviour of a ferrite specimen, on which the amplifier depends for its working, is due not to the polarizing field being such as to give ferromagnetic (Polder) resonance, but to the high concentration of energy in the specimen.

A weakness of the amplifier is that the signal, not being at Kittel resonance, is not concentrated in the specimen and so is not strongly coupled to the pump. If a very large pump power can be generated, so that there is still a large pump power in the ferrite even away from the Kittel condition, it might be possible to improve the performance by arranging for the signal to be in the Kittel condition.

Mr. J. Dain: I should like to ask Prof. Cullen about the phenomenon of gain saturation which he describes as taking place in the travelling-wave parametric amplifier.

Gain saturation appears not to take place in the quadrupole electron-beam amplifier described by Adler. Provided that the electron orbits are not increased in diameter so much that electrons are intercepted by the electrodes of the tube, the gain in the quadrupole amplifying section increases indefinitely with increasing pump power.

In the quadrupole amplifier the phase velocity of the wave is infinite, while in the travelling-wave parametric amplifier it is finite; is this the clue to the difference in behaviour?

Mr. C. R. Russell: In the curve of output power shown by Messrs. Jurkus and Robson, I was rather surprised to see that the idler power was larger than signal power. Could they say something about this, particularly from the physical point of view?

Mr. J. F. Hyde (*communicated*): In the literature on parametric amplifiers confusion is often caused because of loose definitions of terms. This applies particularly to the definition of capacitance. Consider the typical charge/voltage relationship of Fig. E such

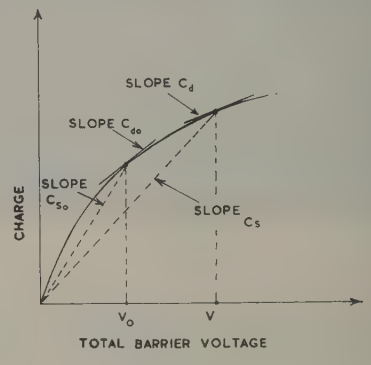


Fig. E

as arises for a semiconductor diode 'varactor'. Here Q is the total unneutralized charge in the high-resistivity side of the depletion layer and V is the total barrier voltage. In the analysis of parametric amplifiers based on such devices we are primarily concerned with the rate of change of this charge with time, i.e. with $i = dQ/dt$. The analysis can proceed in two ways. In the first we write

$$i = \frac{dQ}{dt} = \frac{dQ}{dV}\frac{dV}{dt} = C_d(V)\frac{dV}{dt} \quad . \quad . \quad . \quad \text{(A)}$$

Here, $C_d(V)$ is the dynamic capacitance, which is defined as dQ/dV at V. Alternatively, we may consider the problem in terms of the static capacitance $C_s = Q/V$ at V. Then

$$i = \frac{dQ}{dt} = \frac{d(C_sV)}{dt} = C_s\frac{dV}{dt} + V\frac{dC_s}{dt} = \left(C_s + V\frac{dC_s}{dV}\right)\frac{dV}{dt} \quad \text{(B)}$$

It will be noted that these two methods are equivalent and that

$$C_d = C_s + V\frac{dC_s}{dV} \quad . \quad . \quad . \quad . \quad . \quad (C)$$

Clearly it is simpler to deal with the dynamic capacitance directly and to analyse in terms of a prescribed variation of C_d about the value C_{d0} at the working point defined by V_0, say. For a limited excursion of voltage $V - V_0$ about V_0 we may write

$$C_d = C_{d0}[1 + \alpha(V - V_0)] \quad . \quad . \quad . \quad . \quad (D)$$

As well as yielding simpler analysis, the use of dynamic capacitance is also desirable because it is the dynamic capacitance which we are readily able to measure.

In eqn. (3) of Prof. Cullen's paper, $d(Cv)/dt$ is equated to the change of current per unit length along the line. We must therefore interpret v and C as being the total instantaneous voltage across the line and the instantaneous static capacitance respectively. In his eqn. (1), however, v must be interpreted as the difference between the instantaneous total voltage across the line and the working-point voltage, V_0. These equations are not, therefore, mutually consistent. The difficulty may readily be resolved by writing his eqn. (3) as

$$-\frac{\partial i}{\partial x} = C_d\frac{\partial v}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (E)$$

where $\partial v/\partial t$ is simply the rate of change of voltage across the line, and by defining the dynamic capacitance as

$$C_d = C_{d0}(1 + 2\eta v) \quad . \quad . \quad . \quad . \quad . \quad (F)$$

His eqn. (5) is then unchanged. We note that the wave velocity, u, is equal to $\sqrt{(L_0C_{d0})}$, i.e. it is specified in terms of dynamic inductance and capacitance.

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Professor A. L. Cullen and **Messrs A. Jurkus** and **P. N. Robson** (*in reply*): Mr. White's remarks on the far-reaching simplifications adopted in the analysis are certainly justified. On the other hand, it must be noted that some simplifications are necessary in order to achieve any theoretical treatment at all.

The first point raised by Mr. White has already been partly met by the publication of two papers dealing with chains of identical quadrupoles loaded with diodes,[A,B] though these papers both assume lossless quadrupoles and small-signal conditions.

It is true that lack of uniformity of the diodes presents a very serious problem in experimental work at present. One hopes that the position will improve in this respect and that it will ultimately be possible for manufacturers to supply diodes having a much closer tolerance than at present.

The linear voltage/capacitance characteristic is certainly a fair approximation over a small range of variation, and it can probably be assumed that higher-order terms will bring only a quantitative correction without changing the qualitative aspect of the interaction, at least when only three frequencies are involved.

Experimental work by Engelbrecht[C] and Lombardo and Sard[D] has shown that travelling-wave parametric amplifiers can be made to work. However, it would be premature to answer all the questions raised by Mr. White in his final paragraph, particularly with the adverb 'conclusively'.

In reply to Mr. Dain, it should be noted that the gain of the travelling-wave parametric amplifier can be increased by increasing the pump power, though for a given length of line the gain will not increase monotonically with pump power. However, there is a practical limitation on the maximum gain, which is set by the useful range of voltage of the diodes. This is somewhat akin to the electron interception effect in the Adler tube[E].

The saturation effect referred to by Mr. Dain occurs in the travelling-wave parametric amplifier if the pump is held constant and the input-signal power is increased. In the Adler tube it is obviously true, if r.f. magnetic fields are ignored, that the output-signal power cannot exceed the sum of the input-signal and pump powers. The usual treatment of the theory ignores the pumping power, but assumes a constant pumping voltage. In fact, the transverse motion of the electrons induces a current in the quadrupole structure, and the pumping power is finite. Therefore as the input signal power increases, the pumping voltage must fall if the pump power remains constant. Hence a saturation effect occurs, whether the phase velocity is infinite or not. However, the character of this saturation effect is, of course, different in detail from that described in our papers.

Referring to Mr. Russell's question, the maximum power at the idling frequency is given by

$$P_{2max} = \frac{\omega_2}{\omega_3}P_{30}$$

and the maximum power at the signal frequency by

$$P_{1max} = P_{10} + \frac{\omega_1}{\omega_3}P_{30}$$

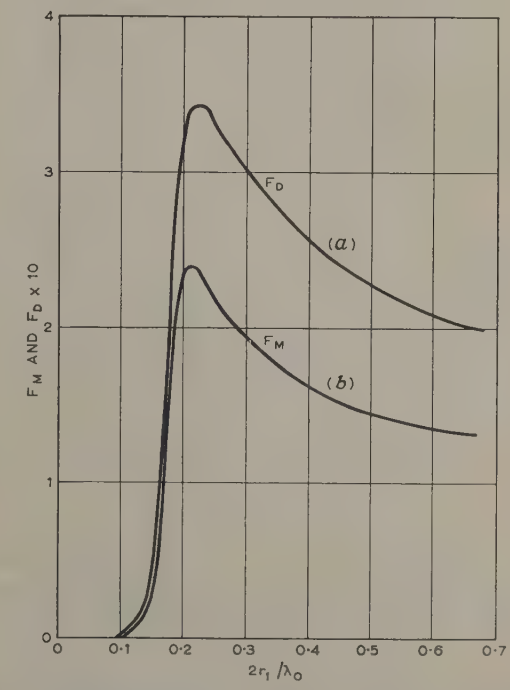


Fig. F.—Attenuation factors for H_{11}-type mode propagating in a circular waveguide of diameter such that $2r_0/\lambda_0 = 0{\cdot}685$, containing a rod with $\mu'/\mu_0 = 1{\cdot}6$, $\epsilon'/\epsilon_0 = 11{\cdot}4$.

Attenuation coefficient $\alpha = \frac{2\pi}{\lambda_0}\left(\frac{\epsilon''}{\epsilon_0}F_D + \frac{\mu''}{\mu_0}F_M\right)$ 8·686 dB/unit length

The relative magnitudes of $P_{1\,max}$ and $P_{2\,max}$ are thus determined by the initial conditions and the frequencies.

We are grateful to Mr. Hyde for drawing our attention to the significance of the bias voltage in dealing with parametric amplifiers using practical variable-capacitance diodes. We entirely agree that the best way to allow for this is to use the dynamic rather than the static capacitance, and we accept his suggestion that the law of the capacitance variation should be eqn. (F); eqn. (E) should then be substituted for eqn. (3) in Prof. Cullen's paper.

Dr. P. J. B. Clarricoats (*in reply*): Mr. Waldron points out correctly that in the case of a ferrite with certain small dimensions, e.g. a thin rod, the signal is not concentrated within the ferrite; thus the amplifier gain is rather low. However, by increasing the diameter of the rod, there is reason to expect that the signal amplification will increase in much the same way as signal attenuation increases when the pump is absent. Fig. F shows the dependence of attenuation on rod diameter for an isotropic rod possessing (*a*) dielectric loss and (*b*) magnetic loss. The curves are normalized so that attenuation can be determined for any value of ϵ'' and μ'', provided only that these values are not so large as to cause the attenuation coefficient, α, to approach the value of the phase coefficient, β. The difference in the attenuation scales arises from the relatively large value of ϵ'/ϵ_0 compared with μ'/μ_0. It is of interest to note that an increase in normalized rod diameter from the value $2r_1/\lambda_0 = 0{\cdot}1$ (above which the perturbation method for determining gain becomes increasingly invalid) to $2r_1/\lambda_0 = 0{\cdot}2$, causes the attenuation factor, F_M, to increase from about 0·002 to 2·0. Thus, significantly greater amplification may be expected from an amplifier using a rod of larger diameter than that employed in the perturbation calculation described in my paper.

With regard to Mr. Waldron's last suggestion relating to the use of the pump in a non-resonant condition, calculation reveals that prohibitively large amounts of pump power would be required for appreciable gain. Even with the pump frequency adjusted to coincide with the Kittel absorption resonance, the large pump power required sets a limitation on the practicability of this type of amplifier. In this respect Mr. Thomson's proposal appears more promising.

References

(A) Kurokawa, K., and Hamasaki, J.: 'Mode Theory of Lossless Periodically Distributed Parametric Amplifiers', *Transactions of the Institute of Radio Engineers*, 1959, MTT–7, p. 360.

(B) Heilmeier, G.: 'An analysis of Parametric Amplification in Periodically Loaded Transmission Lines', *RCA Review*, 1959, **20**, p. 442.

(C) Engelbrecht, R. S.: 'Nonlinear-Reactance (Parametric) Travelling-Wave Amplifiers for UHF', Solid-State Circuits Conference, 1959, digest of technical papers (Lewis Winner, Philadelphia).

(D) Lombardo, P. P., and Sard, E. W.: 'Low-Frequency Prototype Travelling-Wave Reactance Amplifier', *Proceedings of the Institute of Radio Engineers*, 1959, **47**, p. 995.

(E) Adler, R., Hrbek, G., and Wade, G.: 'The Quadrupole Amplifier, a low-Noise Parametric Device', *ibid.*, p. 1713.

621.391.812.63

The Institution of Electrical Engineers
Paper No. 3156 E
Mar. 1960

THE PROPAGATION OF HIGH-FREQUENCY RADIO WAVES TO LONG DISTANCES

By F. KIFT, Associate Member.

(*The paper was first received 2nd June, and in revised form 2nd September*, 1959.)

SUMMARY

The parabolic-layer transmission equation of Appleton and Beynon has been solved for a wide range of its main parameters. A method of applying the equation to any long-distance circuit, in which the successive hop lengths of each transmitted ray are adjusted to accord with ionospheric variations along the path, is described. The only data required are directly obtainable from published ionospheric predictions. The results are displayed on mode plots, which comprise (*a*) a mode-angle plot, showing the angles of elevation of all rays and the modes in which they are propagated to all points along a great circle of any extent, and (*b*) a mode-delay plot, showing the time of travel of all rays instead of their angles of elevation. The effect on the mode plots of diurnal and longer-term changes in the ionosphere is discussed. Two long paths are examined experimentally and it is shown that the number of signal components observed, their relative time-delays, their angles of elevation and their response to varying ionospheric conditions, all agree well with those determined theoretically from the mode plots, provided that the effect of the tropical-Es layer is included. Mode-angle plots may also be used without modification to find both the angular spectrum of back-scatter patterns and the most probable paths taken by atmospheric noise energy arriving at a receiver. Mode-delay plots also provide direct back-scatter patterns. The information on the plots may be used to improve aerial design and also to choose an optimum signalling frequency such that the number of active modes may be reduced, with a possible improvement in delay distortion of the signal. The choice of an optimum signal frequency is made possible by means of a mode-time plot, which shows the diurnal changes to be expected in propagation characteristics as determined by predicted changes in the ionosphere over the path.

LIST OF PRINCIPAL SYMBOLS

E, F1, F2 = Main thick ionospheric layers.
Es = Sporadic-E layer.
D = Spherical (ground-range) distance from transmitter, km.
D_1, D_2 = Part ranges.
P' = Slant range: equivalent path length, km.
f_s = Signal frequency.
f_0 = Ordinary-ray critical frequency.
f_{Es} = Minimum Es penetration frequency of vertically incident waves.
$x = f_s/f_0$ = Frequency factor.
h_m = Height above earth of maximum ion density of a parabolic layer.
h_0 = Height of lower boundary of a parabolic layer.
$y_m = h_m - h_0$ = Semi-thickness of a parabolic layer.
i_0 = Angle of incidence of a ray on a layer at height h_0, rad.
δ = Angle of elevation of a ray at the ground, rad.
δ_{max} = Maximum angle of elevation of Pedersen rays, i.e. the critical angle.
δ_{skip} = Maximum angle of elevation of normal rays, i.e. the skip angle.
$n = \sin i_0 = (R/r_0)\cos\delta$.
3F2, 2Es = Multiple-hop modes. The number of consecutive hops precedes the layer designation.
R = Radius of the earth (6370 km).
$r_0 = R + h_0$.
R = Receiving aerial.
T = Transmitting aerial.
N = Electron density, electrons/cm^3.
t = Propagation time, millisec.
t_0 = Theoretical minimum of t.
$\Delta t = t - t_0$ = Excess time.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

(1) INTRODUCTION

As early as 1938, Millington,[1] in discussing Eckersley's work[2] on the oblique propagation of high-frequency radio waves by way of the ionosphere, pointed out that long-distance paths would consist of unequal hops and that a knowledge of the distribution of ionization at every point along the route was desirable in solving the problem of how energy travels from one point on the earth's surface to another. The object of this paper is to give an approximate solution to this problem for any given state of the ionosphere over the path, and to present the solution in a form readily applicable to any long-distance circuit. The two main simplifying assumptions are that a parabolic relation between height and electron density holds in all thick ionospheric layers and that the angle of elevation of a ray at take-off is equal to its angle of elevation upon arrival at a point, however distant. The treatment employed is based on the parabolic-layer transmission equations of Appleton and Beynon[3] and uses ray theory throughout. Emphasis is placed on the paths followed by the rays between the transmitter and the receiver rather than on the strength of the resulting field, which cannot normally be calculated until these paths are known.

The paper has been written primarily with the needs of the communication planning engineer and the aerial designer in mind. Some of their problems to which it proposes an approximate answer are: (*a*) the choice of an optimum signalling frequency to ensure the longest operating time consistent with low signal distortion, (*b*) the design of aerials which will radiate economically despite the variability of the ionosphere, or which will have maximum pick-up of wanted signals and minimum pick-up of atmospheric noise or interfering signals, (*c*) the causes of circuit failures, and (*d*) the understanding of the propagation processes likely to apply along a given path at a given time.

The effect of the earth's magnetic field has been ignored and ordinary-ray propagation considered only. The theory nevertheless appears to give a good first approximation to the observed facts. The day-time ionosphere often departs from the parabolic model, particularly in equatorial regions, but the overall effect of this departure on long-distance propagation is probably small. Some comparisons between the true ionosphere and the assumed model are made in Section 11.

The method currently in general use for finding the length of a hop made by a ray at a given angle of elevation is that described in Reference 4. It is based on a mirror analogy of the ionosphere and requires either a set of vertical-incidence ionograms for all

points along the path, which are never available, or the use of a rule of thumb in which the heights of reflection are supposed invariable at h'_{min}, the minimum virtual height of the layer, a supposition known to be untrue. A number of approximations are used in correcting from vertical to oblique transmission conditions, and the assumed law of electron distribution is not formulated.[5]

The present method uses the mirror analogy only when rays are propagated by way of the thin sporadic-E layer. For all thick layers the true path of a ray through a parabolic layer is used to determine its hop length. The effect on the ray paths of large layer tilts, which may give abnormal hop lengths, of auroral reflecting curtains or of deviations from the great-circle path are not considered.

(2) RAY-PROPAGATION CHARACTERISTICS

Figs. 1 and 2 show the symbols used to describe the geometry of layers and hop lengths. In describing the dispersive properties of a layer, the symbol x with a subscript to denote the active layer, e.g. x_{F2}, is used to denote the ratio signal/ordinary-ray-critical frequency; this ratio, f_s/f_0, is called the 'frequency factor' throughout. For a fixed signal frequency, the frequency factor is thus inversely proportional to the square root of the maximum electron density of a layer. For a single hop to a given distance a value of x exists above which the ray will skip. This value, x_{skip}, is the maximum-usable-frequency (m.u.f.) factor for that distance, and is normally the subject of much emphasis. The present treatment is concerned, however, more with propagation when frequency factors are below their maximum, since in practice the m.u.f. is rarely used.

The length of a single hop is the ground range, D, and depends on the angle of elevation, δ, of the ray, the frequency factor and the geometry of the layer as described by h_m, h_0, and y_m. Examples of the notation for multiple hops are: (*a*) 2F2 + 1Es + 1F2, which describes the sequence of hops, (*b*) an M mode, in which an intermediate reflection occurs from the top of a lower layer, and (*c*) a 'guided' mode (the G mode) in which a ray undergoes more than one consecutive interlayer reflection.

Solution of the Appleton and Beynon equation shows that the main factors controlling the ground range of a single hop are

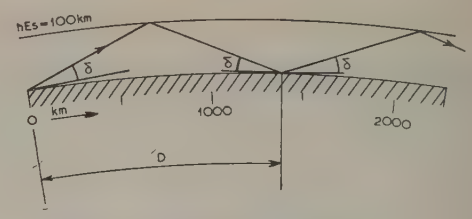


Fig. 1.—Symbols for Es propagation.
(Not to scale.)

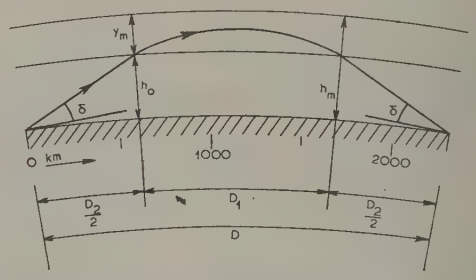


Fig. 2.—Symbols for thick-layer propagation.
(Not to scale.)

δ, x and h_m. The factor y_m/h_0 often has little effect because, for fixed h_m and x, a decrease in y_m causes an increase in h_0, and vice versa; but a decrease in y_m reduces D_1, while the associated rise in h_0 increases D_2. Hence the total ground range D, which is the sum of D_1 and D_2, tends to remain constant.

Fig. 3 shows the changes in the first-hop ground range when h'_m alone varies. To avoid complicating the Figure, the Pedersen (high-angle) rays have been omitted, and only two of the normal rays are shown—the tangent ray ($\delta = 0$) and the skip ray ($\delta = \delta_s$). In practice, the energy downcoming on the illuminated zone will consist of both Pedersen and normal rays. When x is low, Pedersen rays will contribute effectively only to the first few hundred kilometres along the zone, so that at more distant points the normal rays alone will be detectable. When x is high,

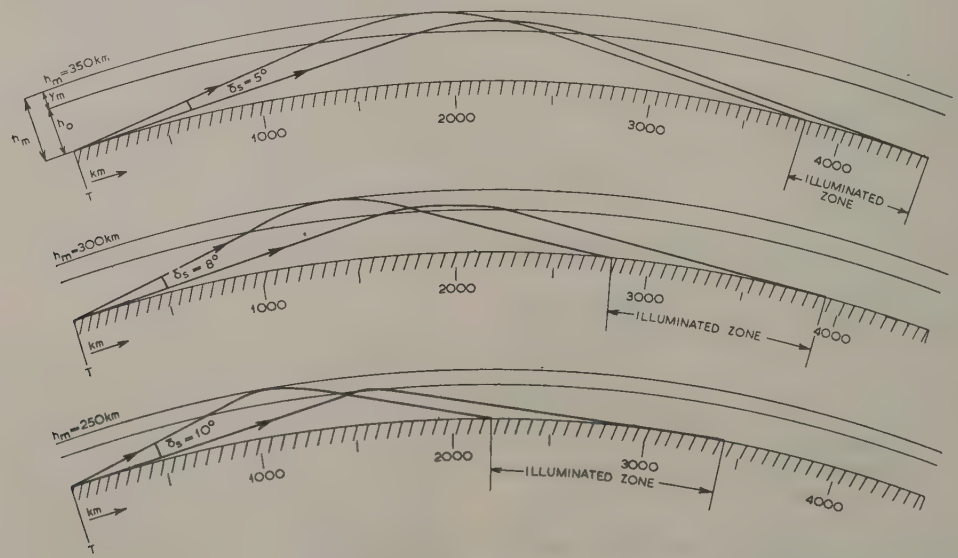


Fig. 3.—Effect of height on propagation.
$x = 3$ $y_m/h_0 = 0{\cdot}4$

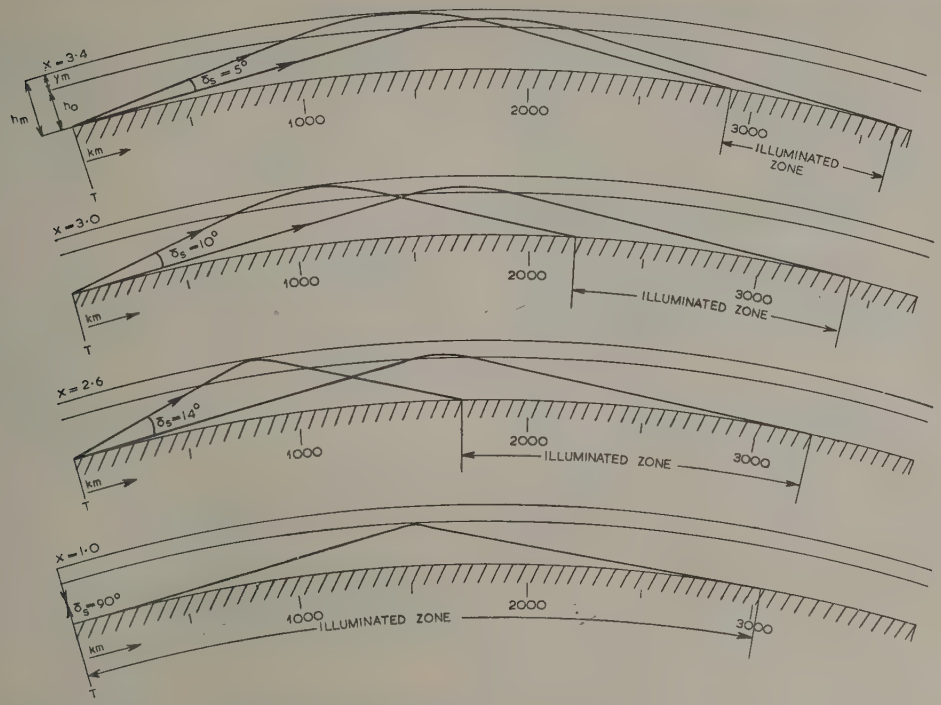


Fig. 4.—Effect of frequency factor on propagation.
$h_m = 250\,\text{km}$ $y_m = 70\,\text{km}$ $y_m/h_0 = 0{\cdot}4$

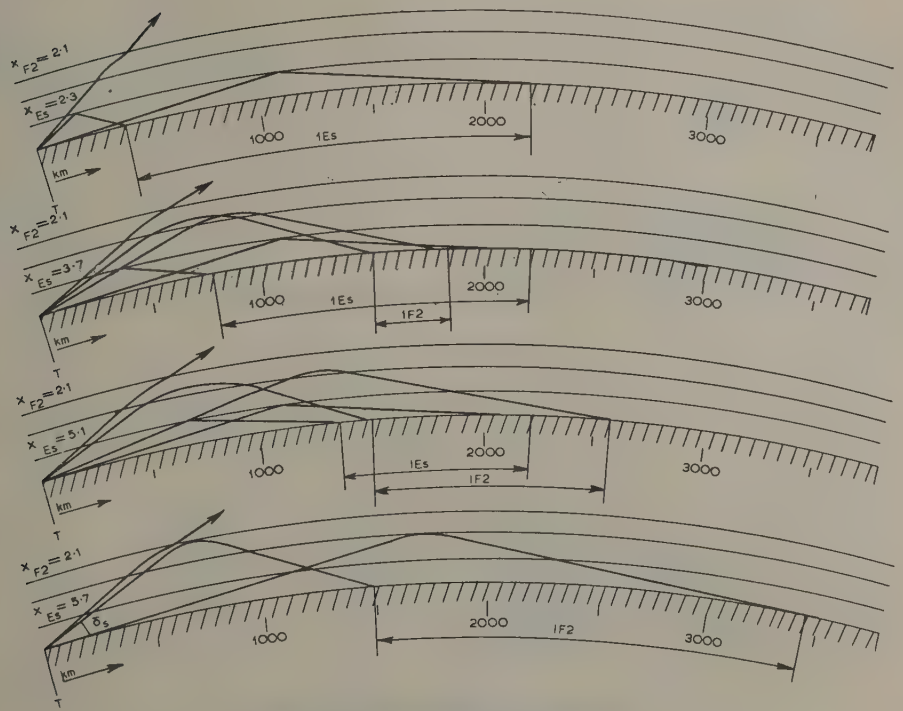


Fig. 5.—Effect of Es layer on propagation.
$h_m = 300\,\text{km}$. $y_m = 85\,\text{km}$. $y_m/h_0 = 0{\cdot}4$

however, both Pedersen and normal rays may contribute to the illumination over the entire zone. As x approaches its maximum, Pedersen rays may be detectable at 5000 km or more,[6] well beyond the single-hop limit of the tangent ray. This condition is discussed in Section 6.1.

Fig. 4 shows the dependence of first-hop ground range on changes in the thick-layer frequency factor. Again, Pedersen rays have been omitted.

Fig. 5 shows how the illumination pattern varies with the frequency factor of a sporadic-E cloud below a thick constant layer.

In practice, both h_m and the frequency factor of each layer may vary from point to point above the great circle, so that the effects illustrated may all operate at once. The method described in Section 3.3 has been developed to enable these variations to be taken into account.

(3) TRANSMISSION EQUATIONS

The basic problem to be solved quantitatively is to find the points to which energy, projected at any frequency and angle of elevation towards a multi-layer ionosphere, will return to earth if the layer characteristics vary with distance away from the source. A point at which a ray returns to earth from the ionosphere will be called its arrival point. The equations used in solving this problem are given below.

(3.1) Parabolic-Layer Equations

The Appleton and Beynon equation for the single-hop ground range, D, of a ray launched at an angle of elevation, δ, towards an ionospheric layer may be written

$$D = D_1 + D_2 = 2R\left\{\frac{x}{r_0}y_m n \operatorname{arc\,tanh}\left[\frac{x\sqrt{(1-n^2)}}{1-x^2y_m n^2/r_0}\right] + \arccos n - \delta\right\} \quad . \quad (1)$$

where $n = (R/r_0)\cos\delta$ and $r_0 = R + h_0$. From eqn. (1), the frequency factor cannot exceed

$$x_{max} = \frac{\sqrt{[1-n^2(1-4y_m/r_0)]} - \sqrt{(1-n^2)}}{2n^2y_m/r_0} \quad . \quad (2)$$

and D_1 then tends to infinity.

Eqn. (1) is plotted in Fig. 6 for an 8° ray and an F2 layer for which $y_m/h_0 = 0{\cdot}4$. The range of the Pedersen rays has been restricted because, although theoretically they extend to infinity in accordance with eqn. (2), they are likely to be very weak at extreme ranges. Further, the longer their range in a varying ionosphere the less is the probability of their hop remaining symmetrical. The broken line represents $D_2[=2R\ (\arccos n - \delta)]$.

Using an automatic digital computer, eqn. (1) has been solved for the thick layers for the range of values shown below.

E layer: $h_m = 110$ and 120 km, $y_m/h_0 = 0{\cdot}2$.
F1 layer: $h_m = 190$ and 220 km, $y_m/h_0 = 0{\cdot}4$.
F2 layer: $h_m = 220 - 500$ km, in 20 km steps, $y_m/h_0 = 0{\cdot}4$ and $0{\cdot}8$.

In all cases $\delta = 0$–60°, and x varies from $x = 0$ to $x = x_{max}$ in suitable steps.

The values of D corresponding to these figures are used in all thick-layer work throughout the paper.

The equation for the equivalent path length, P', for a single hop has been given by Shearman.[7] It may be written

$$P' = D_1\left(\frac{r_0}{nR}\right) + D_2\left(\frac{\sin\theta}{n\theta}\right) \quad . \quad . \quad . \quad (3)$$

where $\theta = \arccos n - \delta$.

The excess path, $P' - D$, is more convenient in practice. It is

$$P' - D = D_1\left(\frac{r_0}{nR} - 1\right) + D_2\left(\frac{\sin\theta}{n\theta} - 1\right) \quad . \quad (4)$$

Eqn. (4) has also been solved for the range of parameters listed above. The excess time, $t - t_0$, is

$$\Delta t = t - t_0 = \frac{P' - D}{c} \quad . \quad . \quad . \quad . \quad (4a)$$

Eqn. (4a) has been used throughout to find the relative delay between signals arriving after different modes of propagation.

(3.2) Thin-Layer (Es) Equations

Putting the thin-layer condition $D_1 = 0$ in eqn. (1),

$$D = 2R\ (\arccos n - \delta) \quad . \quad . \quad . \quad . \quad (5)$$

The upper frequency-factor limit has been assumed to be

$$x_{Es} \gtrdot \frac{1}{\sqrt{(1-n^2)}} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which is equivalent to $x \gtrdot \sec i_0$.

Putting $D_1 = 0$ in eqn. (4) gives the excess time as

$$\Delta t = \frac{P - D}{c} = \frac{D}{c}\left(\frac{\sin\theta}{n\theta} - 1\right) \quad . \quad . \quad . \quad . \quad (7)$$

which has been solved for the same conditions as eqn. (5).

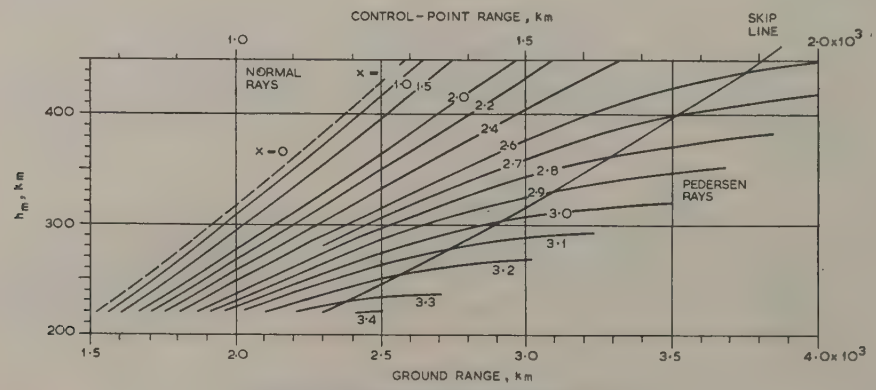


Fig. 6.—Parabolic-layer transmission curves.
$\delta = 8°$. $y_m/h_0 = 0{\cdot}4$

(3.3) Application of Equations

Before the various arrival points of a ray in its progress towards the receiver can be found the state of the ionosphere at the control points of each hop along the great-circle path must be known, and therefore the positions of these control points. One technique we have developed uses a transparency on which are plotted the variations in critical and maximum usable frequencies (or h_m) along the path, as determined from published works.[8,9] This transparency is laid over a family of transmission curves drawn from eqns. (1) and (5) for a particular angle of elevation, i.e. for a particular ray.

For a given signal frequency, the technique automatically gives, for a particular ray: its control point, the ionospheric conditions prevailing there and its arrival point, which is then taken as the launching point for its next hop. The control points are found not by successive approximations but by intersection, so that their accurate positioning depends only on the accuracy of the parabolic or thin-layer models and of the predicted contours. The technique will not be further described here.

Inaccuracies due to errors in predicted contours may be reduced by correcting the transparency with observed data from ionosondes on or near the great-circle path. Inaccuracies due to departures from the parabolic model are discussed in Section 11.

In applying the technique, a theoretical aerial directed along the TR great circle, and radiating equally at all vertical angles from 0–90°, is assumed at the transmitting site, T. The rays leaving T are then dealt with in turn, starting with the tangent ray ($\delta = 0$) and proceeding at regular intervals of δ, the various arrival points of each ray along the great circle being plotted against its angle of elevation until it either penetrates the ionosphere or overshoots the receiving location, R. The effect on each ray of all active layers, including M reflections and guided modes (G modes), is directly apparent using this method. The process is continued until the critical angle for the path is reached above which no ray leaving T is returned to earth.

The resulting sets of arrival points, when joined by smooth curves to include those of the unsampled rays, show not only the extent of all zones illuminated along the TR great circle by the radiator at T but also the active modes of propagation at all points (including R), their associated angles of elevation and the areas of the ionosphere which are determining them.

A similar technique, using eqns. (4) and (7), provides the delay times concerned with each mode of propagation at all points along the great circle.

When each set of arrival points is labelled with its corresponding mode of propagation, the resulting diagram is conveniently termed a mode plot. A mode plot with δ as ordinate will be called a mode angle plot, while a mode plot with Δt as ordinate will be called a mode delay plot. An example of each kind of plot is given in Fig. 7, which will be discussed in more detail.

(4) MODE PLOTS

Referring to the mode plots of Fig. 7, the following is a brief summary of the information yielded by this method of presentation.

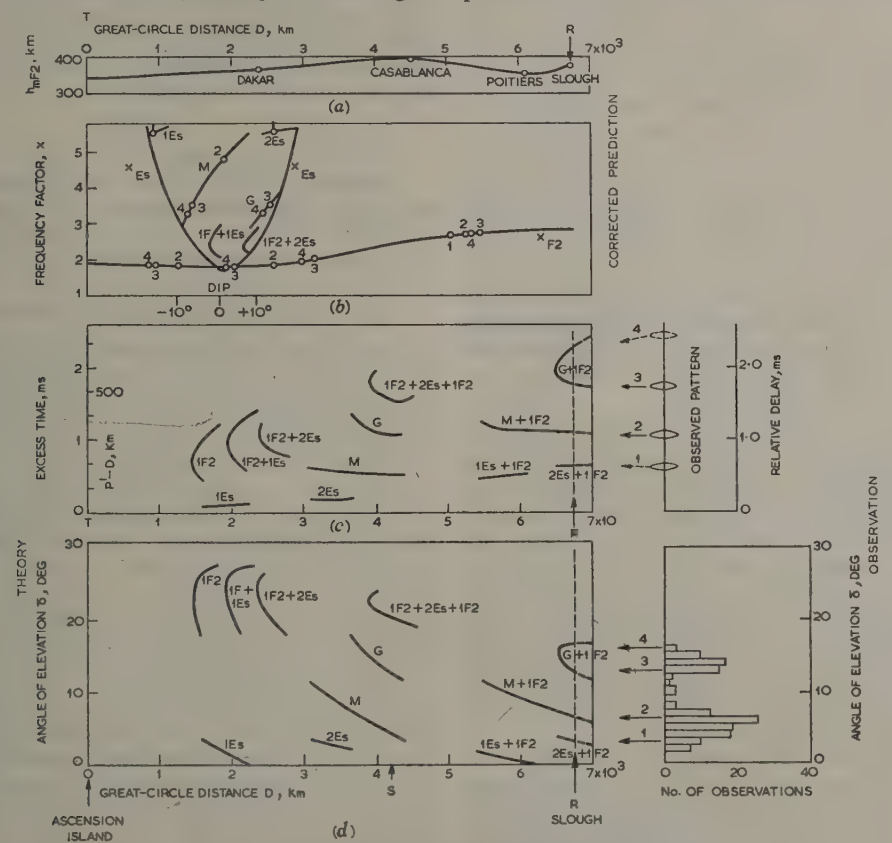


Fig. 7.—19·87 Mc/s mode plots: Ascension Island to Slough.
(*a*) h_m-plot. (*c*) Mode-delay plot.
(*b*) *x*-plot. (*d*) Mode-angle plot.
17th April, 1956, 0930–1020 U.T.

(4.1) Mode-Angle Plots

A mode-angle plot applies to a particular state of the ionosphere above a great circle at a given instant of time. It shows the angular distribution of energy along the path.

The ionospheric conditions prevailing above the Ascension Island–Slough great circle, at the date and time for which the plots were made, are shown at Fig. 7(*a*) and (*b*). These curves were based on predicted characteristics but corrected with observed data from ionosondes at the four points shown (Section 5.1). The curve marked x_{F2} is simply a plot of the F2-layer frequency factor, i.e. $19{\cdot}87/f_{0F2}$. The curve marked x_{Es} is similarly a plot of $19{\cdot}87/f_{Es}$, where f_{Es} is the predicted minimum frequency at which vertically incident energy will penetrate the tropical-Es layer.

The lines marked 1Es, M, G, etc., within the x_{Es} curve define the great-circle extent of the tropical-Es layer zones illuminated by energy propagated in those modes, i.e. the great-circle extent of the 1Es, M, G, etc., control points. At each control point there exists a maximum value of the Es frequency factor above which the mode will fail, and it is thus convenient to plot the great-circle distance of each control point against the corresponding maximum frequency factor, thereby obtaining an estimate of the reliability of each mode. For example, the Es control point of the 5° ray, which is shown by the mode-angle plot to be propagated in the M-mode to about 4000 km, is a point in the Es layer vertically above the earth at 2000 km from Ascension Island. The M line at 2000 km shows that the Es frequency factor would have to exceed about 5 before the 5° ray propagated in this mode would fail. Since the predicted x_{Es} at this point is seen to be less than 2·0, the 5° ray at 4000 km may be considered extremely stable in the sense that it would require a highly improbable variation of x_{Es} from its predicted value to cause the 5° M mode to fail. The reliability of the tropical-Es predictions, and the position of the tropical-Es layer, are discussed in Section 5.1.1.

Each ray reaching Slough (6750 km) has been numbered, the same number being shown against its various control points along the route so that the controlling areas of the ionosphere can be seen at a glance. Further features of Fig. 7 are discussed in Section 5.1.

The angles of elevation and modes of propagation to be expected at Slough (or at any other point along the great circle) when a transmitter is at the origin, can be read directly from the mode-angle plot, and this is its main purpose. Other uses are suggested by the following considerations.

An ordinate erected at a point S (Fig. 7) on the mode-angle plot intersects the mode lines to show the angles of elevation and modes of propagation of all rays arriving at that point from a source at the origin. But some of this arriving energy is returned from S to the origin by back-scattering. It is evident from reciprocity that the back-scattered energy from S can return to the origin only by reversing its original routes. This argument applies to all points along the path, so that the mode-angle plot as it stands gives the complete vertical-angle spectrum and modes of propagation appropriate to a back-scatter pattern from this direction as observed at the origin. To be of use in interpreting an actual back-scatter pattern, mode-angle plots may be needed for a number of great circles covering the azimuthal extent of the radiation from the transmitting aerial.

Apart from the back-scattering sources along the great circle there are also atmospheric-noise sources, to which a similar argument applies. The mode-angle plot thus gives the vertical-angle spectrum and modes of propagation of the signal-frequency component of all possible atmospheric-noise energy arising along the path and entering a receiver situated at the origin (except for local noise). A possible way in which this information could be used to improve signal/noise ratio when the location of a zone of frequent noise is known is described in Section 6.4.

Below the ionosphere, the apex angle of the cone of direct entry of the signal-frequency component of cosmic noise arriving in the neighbourhood of a receiver at the origin is approximately $\pi - 2\delta_{max}$, where δ_{max} is the highest angle on the plot. This is so because the range of angles at which downcoming energy can arrive is by inference identical with the range at which upgoing energy can escape.

(4.2) Mode-Delay Plots

A mode-delay plot, which again applies to a particular instant of time, shows the delay distribution of energy along the path. 'Delay' (Δt) is here defined as time in excess of D/c in accordance with eqns. (4*a*) and (7), so that the actual path time is $\Delta t + D/c$.

Its main purpose is to show the time spread of signals arriving at a given receiver from a transmitter at the origin, so that both the extent of delay distortion for any allotted signal frequency, and the factors controlling it, are known.

The mode-delay plot also gives directly the back-scatter pattern as observed at the origin from the great-circle direction. The slant range of ground back-scatter is simply $D + c\Delta t$, i.e. the sum of the abscissa and ordinate at any point. For back-scatter use, $c\Delta t$ ($= P' - D$) is given on the right-hand side of the Δt scale.

Except for research or circuit failure investigations, the user is more concerned with diurnal or longer-term changes in the angles and time spreads of energy arriving at his receiver than with conditions at a particular instant along the path. Nevertheless, it is only from these basic plots that such information can be reliably obtained. Before this use of the plots is discussed, the experimental evidence for their reliability will be examined.

(5) EXPERIMENTAL RESULTS

Two main sets of data have been obtained and examined, using (*a*) Ascension Island–Slough pulse transmissions, intended to test the validity of the treatment described, and (*b*) Colombo–Slough pulse transmissions undertaken for another purpose.[10]

(5.1) Results of Ascension Island–Slough Pulse Tests

A series of pulse transmissions began in 1956 from a transmitter in Ascension Island (T). At Slough (R) the angles of elevation of the arriving-signal components were measured using standard method and equipment.[11] Transmission was usually on a frequency of 19·870 Mc/s and lasted for one hour, mostly in the morning. A frequency of 21·583 Mc/s was occasionally used concurrently. On 35 days between February and October, 1956, a total of nearly 5000 measurements was made, an average of 140 observations per hour. Many photographs of the time spread of the pulse patterns were also taken. The unit period of measurement was 5 min.

In order to compare measurements with theory, the predicted frequency factors and heights of maximum ion density of the F2 layer along the path were corrected by hourly data from the ionosondes at Dakar (2400 km from Ascension Island), Casablanca (4500 km), Poitiers (6200) and Slough (6750), all of which are on, or near, the great circle. Measured values of f_{Es} at these stations were not used, for reasons discussed later in this Section. Using the corrected x- and h_m-contours, as shown for example in Figs. 7–9, (*a*) and (*b*), mode-angle plots were made for each observing hour to find the theoretical modes of propagation and angles of elevation. A histogram of the observed angles of elevation was then compared for each hour with the theoretical angles on the plots. These histograms are

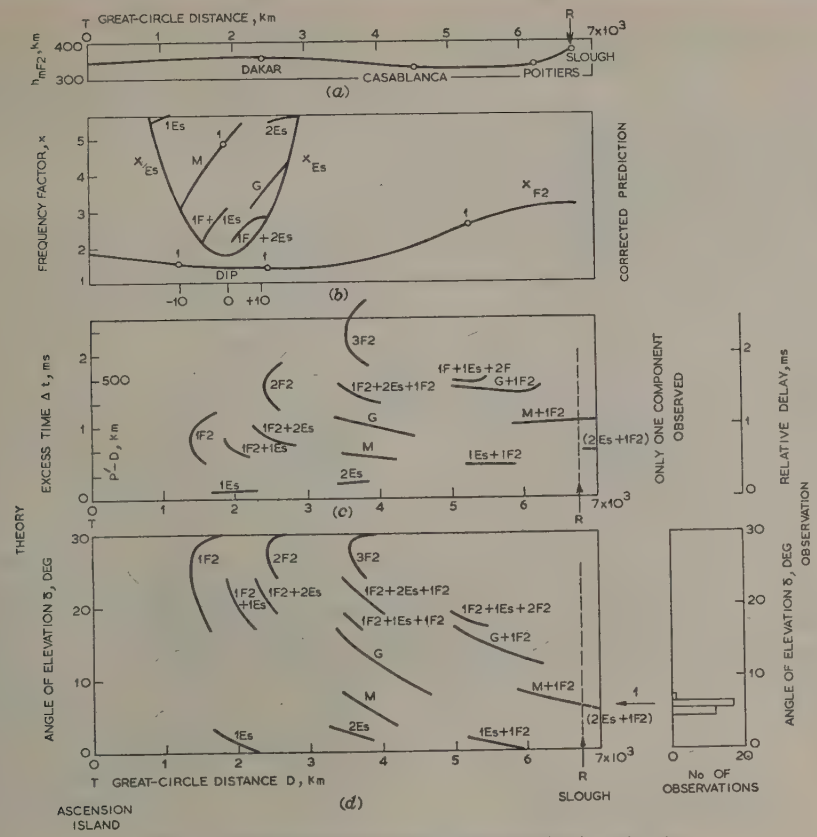


Fig. 8.—19·87 Mc/s mode plots: Ascension Island to Slough.

(*a*) h_m-plot. (*c*) Mode-delay plot.
(*b*) *x*-plot. (*d*) Mode-angle plot.
4th April, 1956, 0940–1030 U.T.

shown to the right of each mode-angle plot. Some mode-delay plots have also been made and compared with observed pulse patterns and delays, and are shown to the right of the mode-delay plots.

Sufficiently close agreement was found between theoretical and observed angles and patterns to justify the use of the parabolic transmission theory outlined. In no period of observations was there any marked deviation from the angles of the corresponding mode-angle plots. The agreement illustrated in Figs. 7, 8 and 9 is typical of the results.

The lowest electron-density areas of the ionosphere along the path, which theoretically determine the maximum angle of a received ray, were always towards the Slough end of the circuit. Figs. 7–9 have been chosen to illustrate conditions of medium, low and high F2-layer electron density, respectively, towards this end of the circuit. E and F1 layers were not active. It is apparent that the main effect of a decrease in ion density is to exclude the higher-angle modes, in agreement with theory.

This dependence of the maximum angle of a received ray on frequency factor and h_m at the low ion-density controlling zone of the ionosphere was found to be very close: in only 3% of the 400 5 min periods of observation did a recorded angle exceed the maximum angle determined by theory from the relevant mode plot.

(5.1.1) Discussion.

Figs. 7–9 show that the measurements concerned only sample a very complex pattern of illumination at one point. To be completely convincing, simultaneous measurements would be needed at other points along the great circle. However, if agreement is found between theory and observation at a single point with respect to (*a*) the number of components, (*b*) their angles of elevation, (*c*) their relative delay pattern, and (*d*) their variations under changing ionospheric conditions, the probability of a correct interpretation is strong. Taking Figs. 7–9 as examples, it will be shown that (*a*)–(*d*) are satisfied.

The complexity of the illumination patterns illustrated is obviously due to the presence of the tropical-Es layer, which has been centred on the dip equator, using predicted[9] median f_{Es}; the density falls off rapidly on either side of the $\pm 10°$ isoclines[12] (marked on Fig. 7).

A guide to the regularity and predictability of the tropical-Es layer, which is associated with the equatorial electrojet, may be seen from the following example for the month and times concerned in Figs. 7–9. Thus during April, 1956, at Kodaikanal, where the dip is $+2°$, the Es layer was observed at this hour (1000 L.M.T.) on every day of the month, median f_{Es} agreed with prediction, while the standard deviation from the average was only 7%. However, at Dakar, where the dip is about $+20°$, although Es was observed on 25 days and median f_{Es} again agreed with prediction, the standard deviation had risen to 15% and the type of Es was often 'blanketing' to vertically incident rays—a characteristic not shared by the tropical-Es layer;[13] Dakar Es may thus be considered beyond the north edge of the tropical-Es layer and too unreliable for use on the mode plots. This does not of course imply that propaga-

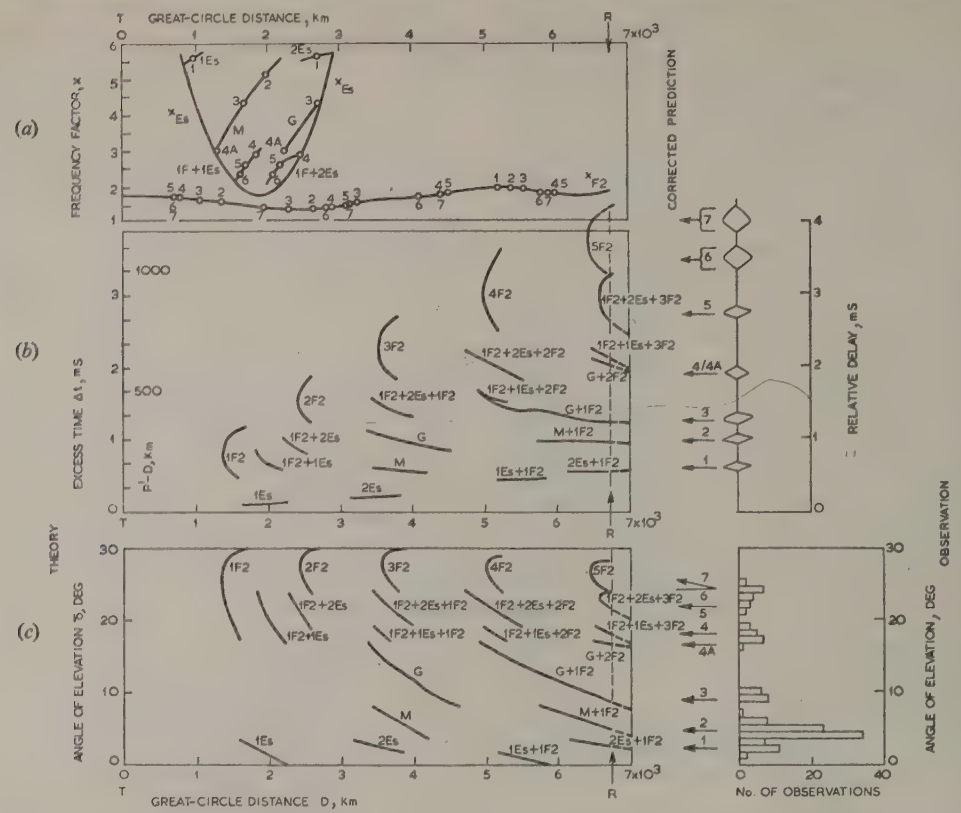


Fig. 9.—19·87 Mc/s mode plots: Ascension Island to Slough.

(*a*) *x*-plot. h_{mF2} fairly constant at 340–360 km.
(*b*) Mode-delay plot.
(*c*) Mode-angle plot.
11th April, 1956, 0930–1030 U.T.

tion is in fact uninfluenced by sporadic clouds forming away from the dip equator, but only that their incidence and density cannot be reliably predicted. The spread of observed low angles on the histograms of Figs. 7–9 is no doubt partly due to these Es clouds. It may be noted that they would have little influence on the path time of the first-arriving components at Slough, which is nearly constant at approximately 23 millisec for low-angle 'mixed' propagation.

Discussing now (*a*)–(*c*) above, comparison of the histograms and pulse patterns with the mode-plots of Figs. 7–9, at 6 750 km, shows that (*a*) the number of signal components agrees with the number predicted, (*b*) the observed angles of elevation agree well with prediction, and (*c*) the relative delay patterns also agree well with the predicted excess time except where Pedersen rays, whose delay is very variable, are concerned. In all these categories pure-F2 propagation would give very poor agreement with observation.

As an aid in assessing the changes to be expected from (*d*), the moment-to-moment variations in the ionosphere, the theoretical positions of the control points of the various components observed at Slough have been marked (i) on the x_{F2} contour and (ii) on the lines within the x_{Es} contour, which are explained in Section 4.1.

On 17th April, Fig. 7, only components 1 and 2 were present at the start of the period of observation (0930 U.T.). At this time ionization (at Slough) was low, so that x_{F2} was higher near this end of the circuit than is shown by the plotted *x*-contour. Towards 1000 U.T. ionization increased to agree with the plotted value and components 3 and 4 were then observed. Later, ionization further increased, component 4 disappeared, component 3 lowered in angle from 13° to 10°, while the relative delay times of components 2 and 3 fell from 1·1 and 1·7 millisec (the values shown) to 0·9 and 1·2 millisec, respectively. Now these are exactly the expected responses of these components to a rising ionization, as may be seen by comparing them with components 2 and 3 of Fig. 9, when the ionization at the Slough end of the circuit was high.

Again, on 4th April, Fig. 8, the low ionization near the Slough end of the circuit would be expected to guide the 2Es + 1F2 rays to a minimum of 6 900 km, i.e. beyond Slough, while the higher rays propagated in the G + 1F2 modes, etc., would penetrate the ionosphere before reaching Slough. Only one mode of propagation, M + 1F2, would therefore be expected, and in fact the rare single-component condition was noted at the time.

On 11th April, Fig. 9, Slough ionization only increased slightly during the observing period. Nevertheless component 3, the G + 1F2 ray, was observed only between 0930 and 0947 U.T., after which time it disappeared. The position of the G control point at the north edge of the tropical-Es layer, marked '3' at approximately 2 800 km, shows that a fall in f_{Es} of only 3–4% would have just this effect, while leaving the other Slough com-

ponents unchanged. Again, during the slight rise in Slough ionization the Pedersen rays, components 6 and 7, became much weaker and eventually failed, in agreement with theory.

This consistency was often evident, and together with the general agreement of observation with theory provides strong evidence that the mode plots give a realistic statement of propagation above the path.

(5.2) Results of Colombo–Slough Pulse Transmissions

Measurements of the angles of elevation of pulses received at Slough from Colombo (8 700 km) on 16·160 Mc/s during daylight hours on one day per week throughout 1953, excluding June to September, have been given statistically elsewhere.[10] A brief analysis of the path conditions, using the mode-plot technique, will be given here.

Hourly ionospheric data for 1953 are available at Slough from Bombay, Madras, Graz, Lindau, De Bilt and Slough, but not from that part of the path over Soviet territories, so that good correction of x- and h_m-contours is not at present possible. However, the predicted x-contours showed that the low-ionization controlling zone would be expected at either the European or the Asian end of the path at the times concerned.

Mode plots were made for one day of each month, and good agreement was found throughout between observed and theoretical angles of elevation and excess delay times. The highest angle of elevation recorded during each of the 490 5 min periods of observation was also compared with the maximum angle to be expected from the low-ionization controlling zones, and it was found that the observed angle exceeded the theoretical angle in only 5% of all the periods of observation, again amply justifying the theory.

(5.2.1) Discussion.

Fig. 10 illustrates conditions between 0930 and 1030 U.T. on the 2nd October, 1953. Because Colombo is only 5° south of the dip equator the general illumination pattern is less broken up, M and G modes being unlikely. Again, pure-F2 propagation—except at high angles—is rare along this path in sunlight. It will be noted that (*a*)–(*c*) of Section 5.1.1 are satisfied. This agreement was general for the hours examined, but was not so striking as for the Ascension Island circuit, probably because less ionospheric data were available for correcting the Colombo path contours.

Agreement under condition (*d*) was difficult to assess, because the lack of hourly Es data during 1953 from ionosondes between the 10° isoclines prevented a check on the reliability of f_{Es} predictions. The tropical-Es layer frequency factor was therefore plotted directly from predictions without checking and centred on the dip equator. Without the introduction of this layer, agreement was again generally poor.

(5.3) Other Evidence

The mode-plot technique has not been confined to the two circuits described, but has been used with consistent success in back-scatter work at Slough and elsewhere, and for the examination of many other circuits. These results will be described elsewhere.

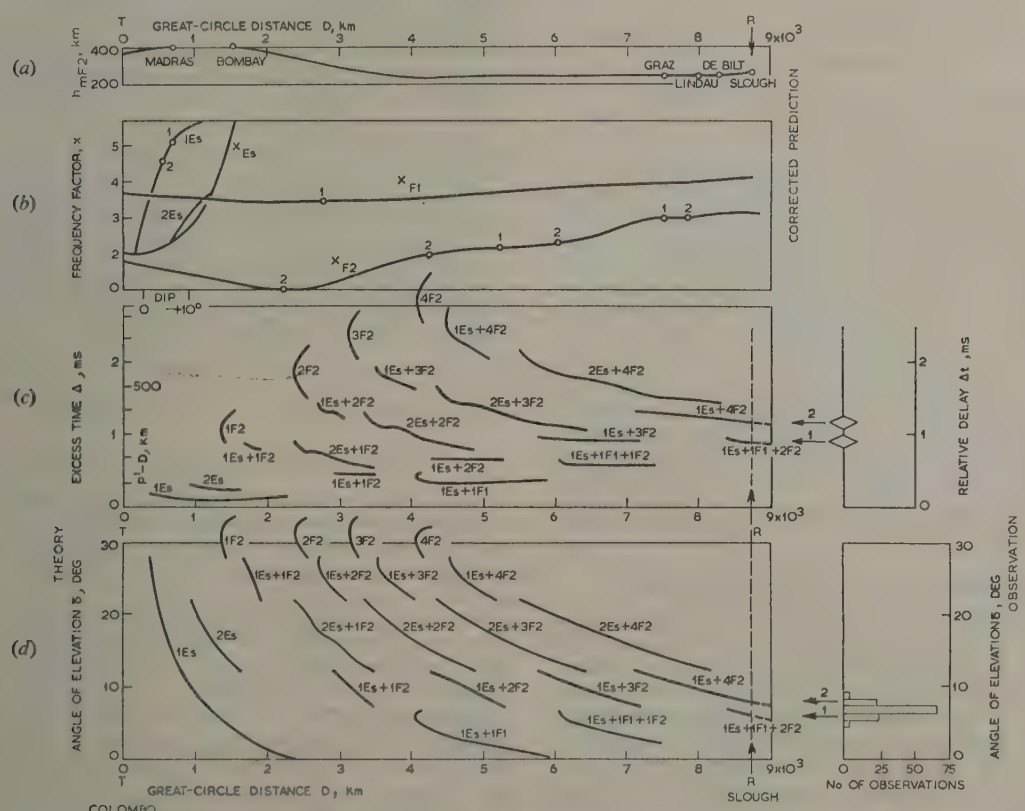


Fig. 10.—16·16 Mc/s mode plots: Colombo to Slough.
(*a*) h_m-plot. (*c*) Mode-delay plot.
(*b*) x-plot. (*d*) Mode-angle plot.

(6) SOME GENERAL FEATURES OF LONG-DISTANCE PROPAGATION

Section 5 showed that propagation may be complex over transtropical paths, and that this complexity is due to the tropical-Es layer. During equatorial darkness, when only residual Es is usually present, or over non-tropical paths, conditions may be much simpler. This Section discusses path illumination during these simpler conditions, as shown by the mode-plot technique, and when only a single thick layer is active.

(6.1) Pedersen Rays

The characteristics of first-hop Pedersen rays are shown in Fig. 11 for fixed h_{mF2} and varying frequency factor, x, at the control points. An arbitrary limit (dashed curve) has been set

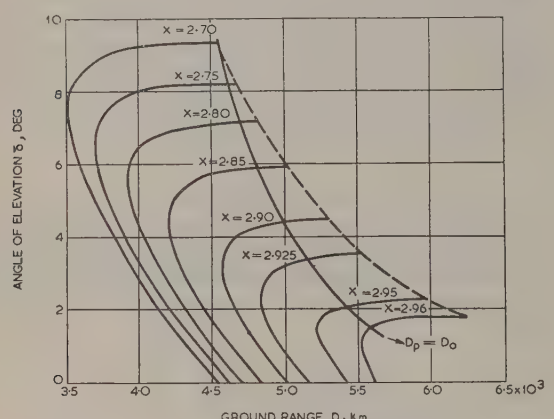


Fig. 11.—Pedersen rays during limiting conditions.
$h_m = 400\,\text{km}$ $y_m/h_0 = 0{\cdot}4$

to their maximum distance, which is theoretically infinite. The curve $D_P = D_0$ defines the angle at which the Pedersen-ray ground range equals the tangent-ray ground range, so that for a further increase in angle of elevation a zone of the earth exists which is illuminated by Pedersen rays alone. This zone lies roughly between the $D_P = D_0$ line and the dashed line. It will be seen that for this value of h_{mF2} (400 km) first-hop Pedersen rays can only exist alone below about 9°. They have been observed quite strongly at 5 300 km,[6, 14] but Fig. 11 shows that they may be unstable at these long distances in the sense that only a small rise in critical frequency at their control points may cause them to fall short of a given distance when a fixed signal frequency is in use. Their angles of elevation and path times are also very frequency dependent. They may survive in subsequent hops only if ionization increases along the path, since they will then be partly refocused.

(6.2) Short-Term Variations: Mode-Stability

The modes of propagation of the rays leaving T may be considered stable if they fail only when abrupt changes occur in critical frequency, such as those associated with sudden ionospheric events or the occurrence of dawn or dusk along the path. Again, for a mode to be considered stable the angle of elevation of the ray arriving at a fixed receiver should not vary by more than 3°–4° (say) despite the normal hour-to-hour, or day-to-day, fluctuations in ionization along the path while a given signal frequency is in use.

In pure-F2 propagation it has been seen that Pedersen rays are not stable in this sense. In mixed propagation it has also been seen that low rays whose control points fringe the tropical-Es layer are unstable. Again, in pure-F2 propagation it is easy to see from mode plots that the lowest-angle rays are those least affected by ionospheric changes, provided that very steep gradients do not occur near the terminals. It is of course possible, by choosing a very high signal frequency, to excite low-angle unstable modes, as illustrated by the 5 000 km point on Fig. 11, but such frequencies are usually avoided in practice. When pure-F2 modes alone are active, and for distances up to about 8 000 km, the mode-plot technique will show the optimum modes and angles of elevation; for distances greater than this the mode plots show that the optimum range of angles of elevation for stable modes is approximately 4°–8°. When transtropical circuits are concerned, it is usually necessary to use the mode-plot technique for any distance, since the effect of tropical Es in breaking up the illumination patterns is very variable and depends on the exact location of the transmitter.

(6.3) Long-Term Variations

When the F2 layer alone is active, the modes of propagation of low rays along any path from which dawn or dusk is excluded can be kept the same by a suitable choice of signal frequency whatever the season of the year or phase of the sunspot cycle. It is of course true that the critical frequency level of the F2 layer varies considerably from season to season and throughout the sunspot cycle and that the shape of the critical-frequency contour over any path may also vary. But it is not the level of critical frequency that determines the modes in which rays are propagated but that of the frequency factor, which can be held relatively constant if a suitable signal frequency is available. Any long-term changes in the shape of the frequency-factor contour will, as shown in Section 5, affect the high-ray modes of propagation in detail, but usually leave the low rays relatively unaffected.

When mixed modes of propagation are active, this generalization fails because the extent and ionization of Es clouds and of residual Es do not follow the sunspot cycle, or necessarily vary seasonally in the same way as those of the thick layers. Thus, in a period of few sunspots, when thick-layer ionization is low and signal frequencies are therefore also kept low in order to ensure regular communication, the level of sporadic ionization may be as high as during a period of many sunspots. The general level of Es frequency factors may then be very low compared with the thick-layer frequency factors, and mixed propagation by way of sporadic clouds or the residual layer may be frequent and, in the present state of knowledge, unpredictable. Even in these circumstances, however, low-ray modes of propagation will usually be more stable than those of the high rays.

(6.4) Propagation of Atmospheric Noise Energy

It was noted in Section 4.1 that the mode-angle plots show all the paths available to atmospheric noise arising along the great circle and entering a receiver situated at the origin. The plots may thus be used to decide the relative merits of transmission at various angles of elevation. For instance, the main high-noise zone affecting the Slough–Ascension Island circuit lies, very approximately, between the points 1 100 and 2 200 km from Ascension Island.[15] Suppose that the optimum major-lobe vertical angle of an aerial with a 6° beam width is required. Then reference to Figs. 7–9 will show that rays projected from Slough at 5° ± 3° will usually arrive at Ascension Island by two signal modes, 1F2 + M and 1F2 + 2Es, often competing with path noise arriving by one mode, 1Es, only, and also that the effective noise zone is reduced to a belt of about 250 km extent. But rays projected at 15° ± 3° would usually reach Ascension Island in one mode only, 1F2 + G or 3F2 + 1Es + 1F2, often

competing with two noise modes, 1Es + 1F2 and 1F2, while the effective extent of the high-noise belt is increased to about 750 km. Some improvement in signal/noise ratio would thus be expected if the aerials were designed to excite the low-angle modes exclusively.

Mode-angle plots constructed for the great circle between an interfering transmitter and the receiver may also be used in a similar way to show if the interference can be reduced by a suitable choice of vertical transmission angle.

(7) MODE-TIME PLOTS AND MAXIMUM-USABLE FREQUENCY

The use of an appropriate signal frequency for long-distance communication is of primary importance. If the signal frequency, and therefore the frequency factor along the path, is too low, much of the radiated energy will be wasted because it illuminates inessential zones of the earth's surface or because it travels to the receiver by an unnecessarily large number of hops and is therefore excessively dissipated in transit. Again, with low frequency factors, many different paths between transmitter and receiver are open to each group of waves constituting a signalling element, and since each path is traversed in a different length of time, any one signalling element arrives at the receiver as a set of confused overlapping groups of waves extensively spread out in time. The modulation waveform at the transmitter is then not reproducible at the receiver.

The use of a too high signal frequency avoids this delay distortion and the wastage of radiated energy on inessential zones of the earth, but is a false economy since not only may the angle of elevation vary greatly but communication may suddenly fail altogether when relatively slight layer changes occur over the path. Again, the radiated energy may be excessively dissipated in its long paths through an active layer.

(7.1) Optimum Working Frequency

The m.u.f. for a given distance is here meant as that frequency above which no signals will be received at that distance. In terms of the frequency factor, and for a single hop, it is $x_{skip}f_0$. The optimum working frequency (o.w.f.) is chosen where possible at 85% of the m.u.f., i.e. at $0{\cdot}85x_{skip}f_0$. This choice assumes that communication is always possible between an absorption-limited lower frequency and the m.u.f. This assumption is demonstrably untrue. Fig. 12, for instance, is a parabolic-layer plot of path

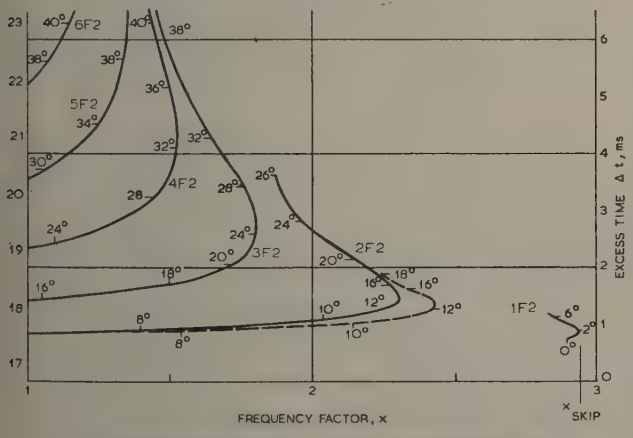


Fig. 12.—F2-layer propagation to 5 000 km.
$h_m = 400$ km $y_n/h_0 = 0{\cdot}4$
---- f_{oF2} at first hop control point increased by 30%.

time against frequency factor, a generalized $P'f$ curve, for a 5 000 km path. It shows the modes of propagation, delay times and angles of elevation of all rays reaching a receiver at 5 000 km. The frequency of transmission is supposed to be swept continuously from the critical frequency of the layer, assumed constant, where $x = 1$, up to the m.u.f. at $x_{skip}f_0 = 2{\cdot}94f_0$. The o.w.f. is at $0{\cdot}85 \times 2{\cdot}94f_0 = 2{\cdot}5f_0$, where communication is impossible. Fig. 11 shows that for this layer the 1F2 Pedersen rays will be undetectable above about 6°, where $x \lessapprox 2{\cdot}84$, while the nose of the 2F2-mode line in Fig. 12 cannot exceed $x = 2{\cdot}3$ for a level layer, or 2·44 (dashed line), even when the critical frequency at its first control point is raised by 30%. It will be apparent that to an observer with a limited number of available frequencies the m.u.f. may well be recorded as less than about $xf_0 = 2{\cdot}3f_0$, since the frequency range in which the 1F2 mode is detectable is very limited. Statistical summaries of the m.u.f. based on commercial usage may thus be very misleading.

Again, Figs. 7–9 show that an observer at 3 000 or 4 700 km will rarely receive any 20 Mc/s signals from an Ascension Island transmitter although x_{F2} is well below the m.u.f. factor.

(7.2) Mode-Time Plots and Optimum Signal Frequency

Possible alternatives to the m.u.f. are the optimum-signal-frequency bands for each mode. These would show the safe frequency ranges in which each mode could be used, bearing in mind delay distortion and stability during normal ionospheric variations, and would include the effect of tropical Es where appropriate. The analysis of a circuit in these, or similar, terms requires data from a number of mode-plots to cover the period for which the circuit has to be kept open, and would most usefully be tackled with an automatic computer. Where the choice of signal frequencies for a circuit is limited, the following simpler method could be used.

It is required to know, for each available signal frequency, the changes to be expected in the angles of elevation and delay times of the rays arriving at a given receiver during the period that the circuit needs to be kept open. The mode plots so far discussed, although applying strictly only to a given instant of time, may in practice show little variation over a one-hour period when ionospheric conditions are relatively stable. For less stable periods, such as at dawn or dusk, the mode plots must, however, be constructed for more frequent intervals.

If the data applying to the receiver location are read from each of these mode plots and assembled on a universal-time instead of a distance axis the required information is available. Such a plot will be called a 'mode-time plot'; examples using three different signal frequencies for the day-time Ascension Island–Slough circuit are shown in Fig. 13. These plots are based on predicted F-2 layer variations only, in order to simplify explanation; the effect on them of the tropical-Es layer will be considered later.

The way in which the circuit opening conditions of Fig. 13(*a*) were built up is shown in Fig. 14. This consists of 8 superimposed mode-angle plots, each labelled with the universal time for which it was constructed, and clearly shows the dynamics of the morning build-up of propagation along the path from dawn onwards. Thus at 0610 U.T. all rays leaving T escape except those in the unstable 1F2 mode beyond 4 000 km. By 0620 U.T. the skip ray in the 2F2 mode, previously more distant, reaches Slough (R) (compare Fig. 13). The 2F2 rays at R then pass through a brief unstable 'focused' stage and by 0700 U.T. are stabilized at R in the 2F2 mode. The 3F2 skip ray reaches R by 0615 U.T., the 3F2 mode becoming stable by 0800. By 1100 U.T. the 4F2 skip ray has reached R and remains relatively unstable, in the sense that both normal and Pedersen rays are present, until about 1300 U.T. (compare Fig. 13). A similar

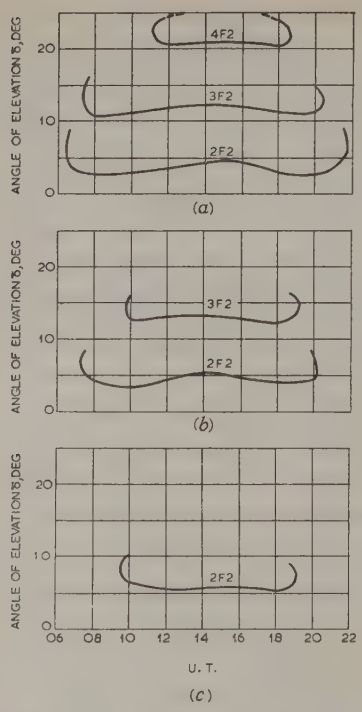


Fig. 13.—Mode-time plots for F2 layer: Ascension Island to Slough.
(*a*) $f_s = 17$ Mc/s. (*b*) $f_s = 20$ Mc/s. (*c*) $f_s = 23$ Mc/s.
April, 1956.

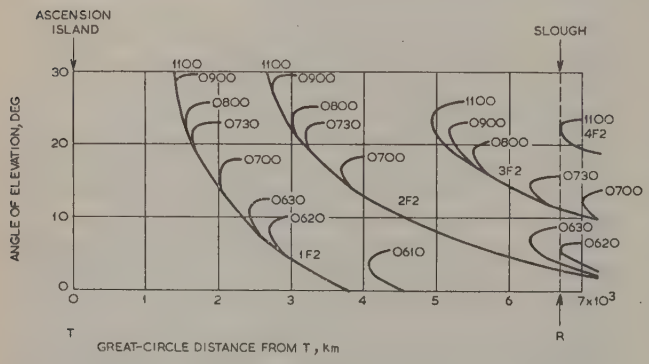


Fig. 14.—17 Mc/s mode-angle plots for circuit opening conditions for F2 layer: Ascension Island to Slough.
April, 1956.

picture holds for the decay of propagation as ionization decreases towards dusk.

Fig. 13(*b*) and (*c*) were similarly constructed using timed mode-angle plots. It will be seen that, by this method, not only can an optimum signal frequency be chosen, from those available, to exclude rays in high-order modes that may be causing delay distortion, while those in a low-order mode can still be utilized, but also valuable insight is obtained into the total effect of such a choice.

Fig. 15 shows a complete diurnal mode-time plot based on predicted F2-layer conditions, when a signal frequency of 20 Mc/s is used during the day and 10 Mc/s during the night. The vertical polar diagram of a fixed simple aerial, supposed to be used for both 10 and 20 Mc/s transmission, is also shown. The failure of all but the low-ray modes at approximately 0600 U.T. refutes the fairly common generalization that angles of elevation are higher at night. No evidence suggests that low rays fail to be propagated by way of night-time layers, but their detection

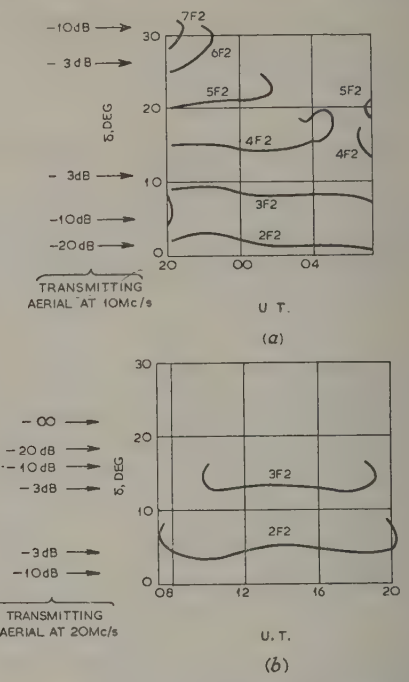


Fig. 15.—Mode time plots for F2 layer only at Slough: Ascension Island to Slough.
(*a*) $f_s = 10$ Mc/s. (*b*) $f_s = 20$ Mc/s.
April, 1956.

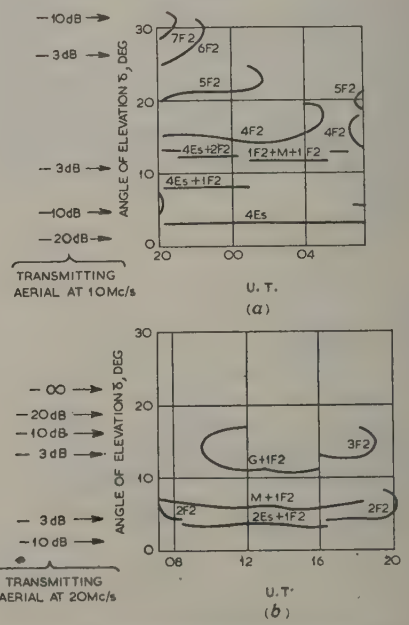


Fig. 16.—Mode-time plots at Slough: Ascension Island to Slough.
(*a*) $f_s = 10$ Mc/s. (*b*) $f_s = 20$ Mc/s.
April, 1956.

may be prevented by the increased angle of elevation of the major lobe if the day-time aerial is used at night, as is shown in Fig. 15.

When the diurnal changes in all layers are taken into account, the result is a complex mode-time plot such as that shown in Fig. 16, where propagation by way of the tropical-Es layer is included. The effect of the predicted median values of f_{Es} in the residual-Es layer during darkness has also been included to show the possibility that, despite its low density, the sporadic layer may still remain active. This may occur because the reduced night-time signal frequency lowers the level of the Es frequency factor. (In practice, residual-Es ionization will never conform to predicted median values over a wide area, so that the night-time plot gives only a qualitative picture of the mixed modes.)

Fig. 16 shows that the choice of an optimum signal frequency may be difficult for a transtropical circuit. The use of the present treatment should nevertheless enable a more informed choice to be made than has previously been possible using two-control-point m.u.f. methods.

Finally, it may be mentioned that all the techniques described in the paper are readily adaptable to the use of an automatic computer.

(8) CONCLUSIONS

The evaluation and application of the Appleton and Beynon expression for the ground range of rays propagated by way of a parabolic layer have shown that the calculated angles of elevation and relative time delays of rays arriving at a distant receiver are close to those observed in practice, provided that each calculated hop length is adjusted to accord with changes in the active layers over the path and also that the propagation of rays by way of the tropical-Es layer is taken into account. The good agreement between theory and observation, for both the Ascension Island–Slough and Colombo–Slough paths, suggests that the theory may be confidently applied more generally. On the assumption that this generalization is justified, many propagation problems can be usefully attacked. In particular, approximate solutions can be given to the problems of the propagation of atmospheric noise energy, the relative stability of different modes of propagation, the design of suitable aerial systems and the changes in propagation characteristics to be expected from long- and short-term changes in the ionosphere. The choice of an optimum signal frequency for use with any long path can be made with more confidence than has previously been possible, since the effect of the total ionosphere over the path is assessable. The use of mode-angle, mode-delay and mode-time plots in choosing an optimum signal frequency should help to lengthen service time and sometimes permit increased signalling speed, since delay distortion can be reduced.

(9) ACKNOWLEDGMENTS

The measurements described in Section 5 were made jointly by Mr. A. F. Wilkins and the author. The author is indebted to Mr. R. J. Hitchcock and to Cable and Wireless Ltd. for valuable co-operation in the Ascension Island–Slough tests. Mr. J. Neale of Government Communications Headquarters kindly gave permission for an automatic digital computer to be used to extend the evaluation of the Appleton and Beynon transmission equations. Capt. Halley of the Centre National d'Études des Télécommunications Section de Prévisions Ionosphériques Nationale made available hourly ionospheric data from Dakar and Casablanca.

The work described was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(10) REFERENCES

(1) Millington, G.: 'The Relation between Ionospheric Transmission Phenomena at Oblique Incidence and those at Vertical Incidence', *Proceedings of the Physical Society*, 1938, **50**, p. 823 and Fig. 8.

(2) Eckersley, T. L.: 'Studies in Radio Transmission', *Journal I.E.E.*, 1932, **71**, p. 405.

(3) Appleton, E. V., and Beynon, W. J. G.: 'The Application of Ionospheric Data to Radio Communication Problems: Part I', *Proceedings of the Physical Society*, 1940, **52**, p. 518.

(4) National Bureau of Standards: 'Ionospheric Radio Propagation', Circular No. 462, June, 1948, p. 67.

(5) Newbern Smith: 'The Relation of Radio Sky-Wave Transmission to Ionosphere Measurements', *Proceedings of the Institute of Radio Engineers*, 1939, **27**, p. 332.

(6) Kift, F.: 'Single-Hop Propagation of Radio Waves to a Distance of 5300 km', *Nature*, 1958, **181**, p. 1459.

(7) Shearman, E. D. R.: 'The Technique of Ionospheric Investigation using Ground Back-Scatter', *Proceedings I.E.E.*, Paper No. 1915 R, October, 1955 (**103** B, p. 210).

(8) 'Predictions of Radio Wave Propagation Conditions', Bulletin A, D.S.I.R. Radio Research Station.

(9) 'Basic Radio Predictions', C.R.P.L., Series D, National Bureau of Standards.

(10) Wilkins, A. F., and Kift, F.: 'Characteristics of H.F. Signals', *Electronic and Radio Engineer*, 1957, **34**, p. 335.

(11) Wilkins, A. F., and Minnis, C. M.: 'Arrival Angle of H.F. Waves', *Wireless Engineer*, 1956, **33**, p. 47.

(12) Smith, Jnr., E. K.: 'A Study of Sporadic E on a World-Wide Basis', National Bureau of Standards Report No. 3575, December, 1955.

(13) Smith, Jnr., E. K.: *ibid*, p. 37.

(14) Warren, E., and Hagg, E. L.: 'Single-Hop Propagation of Radio Waves to a Distance of 5300 km', *Nature*, 1958, **181**, p. 34.

(15) 'Survey of Existing Information and Data on Radio Noise over Frequency Range 1–30 Mc/s', H.M. Stationery Office, Radio Research Special Report No. 15, 1947.

(16) 'Ionospheric N(h) Profiles', D.S.I.R. Radio Research Station.

(17) Vickers, M. D.: 'The Calculation of the M.U.F. Factor for a Non-Parabolic Ionospheric Layer', *Journal of Atmospheric and Terrestrial Physics*, 1959, **17**, p. 40.

(11) APPENDIX

Errors Introduced by the Use of the Parabolic Model

During night-time in most parts of the world the thick ionospheric layer approximates closely to the parabolic model, and errors introduced by using the technique described are probably negligible for practical purposes. It is, however, known from published N(h) data,[16] which accurately describe the ionosphere for a given ionogram, that the day-time ionosphere may be far from the parabolic model deduced from the same ionogram by standard reduction techniques.

Vickers[17] has calculated the ground range of rays travelling by way of the real ionosphere, and two of his results are shown in Figs. 17 and 18, together with the corresponding ground range using the parabolic model deduced from the same ionograms. The examples illustrated are those of extreme departure —in practice much closer agreement would usually be expected. Since present predictions of ionospheric conditions are based on parabolic, or near-parabolic, models as deduced from ionograms, the Figures illustrate the degree of error possible in

extreme cases when the techniques described in this paper are used. If, as a result of N(h) investigations, a better model is proposed, the technique may still be used with new transmission equations.

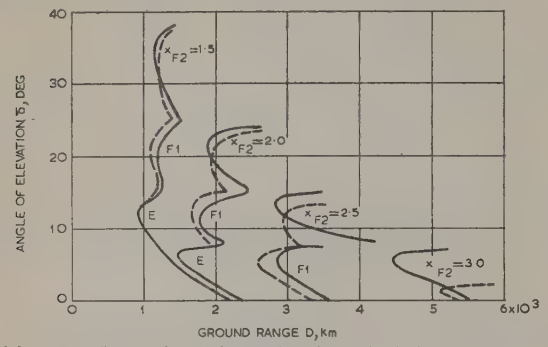


Fig. 17.—Comparison of actual and parabolic ionosphere over Slough.

—— Actual. ----- Parabolic.
$o_{F2}/f_{oF1} = 1{\cdot}4$ $h_{(true)F2} = 325$ km $h_{mF2} = 400$ km $y_m/h_0 = 0{\cdot}4$
27th July, 1957, 1200 U.T.

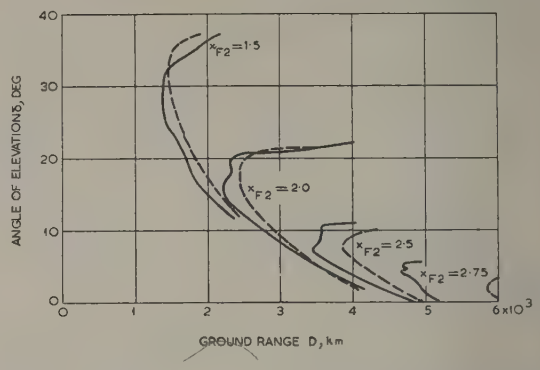


Fig. 18.—Comparison of actual and parabolic ionosphere over Khartoum.

—— Actual. ----- Parabolic.
$h_{(true)F2} = 425$ km $h_{mF2} = 480$ km $y_m/h_0 = 0{\cdot}4$
9th March, 1957, 1300 U.T.

621.397.812 : 621.397

The Institution of Electrical Engineers
Paper No. 3169 E
Mar. 1960
©

V.H.F. FIELD-STRENGTH MEASUREMENTS OVER PATHS IN THE IRISH SEA INVOLVING MOUNTAIN OBSTACLES

Their Application to a Particular Television Channel-Sharing Problem

By J. K. S. JOWETT, B.Sc.(Eng.), Member.

(*The paper was first received 8th July, and in revised form 30th September*, 1959.)

SUMMARY

A description is given of measurements made to compare the propagation of v.h.f. signals over paths in Cardigan Bay and the Irish Sea, some of which involve transmission over high mountain ridges. Comparisons are made, based on simple knife-edge diffraction theory used together with published propagation data, between the expected and measured field strengths at various distances beyond the Isle of Man mountain ridge. The results of the second series of experiments are applied to confirm that interference between co-channel television stations at Winter Hill, Lancashire and Black Mountain, Belfast should be at an acceptable level.

(1) INTRODUCTION

(1.1) General Comments

The characteristics of v.h.f. propagation over long non-optical paths have been the subject of study for a number of years, partly because the successful planning of broadcast and tropospheric-scatter services depends upon a thorough knowledge of this subject. Considerable scientific interest also is centred upon the means by which such signals are propagated, and various explanations of the phenomena involved have been put forward.

The fact that v.h.f. signals are transmitted unexpectedly long distances over sea as well as over land paths, together with other observed features, has led most workers to reject any theory involving only diffraction over uneven ground and to concentrate instead upon tropospheric effects. Nevertheless, as pointed out by Bullington[1] and others, diffraction by terrain may well be a major cause of the propagation of high-power transmissions to intermediate distances of up to 200 km or more. This may apply particularly where the ground is irregular and where no more than one or two large diffracting ridges intervene between the transmitting and receiving points. It is already well known from a number of experiments[2-6] that, in such circumstances, the strength of v.h.f. signals may be greater than it is over paths which contain no large intervening obstacle but which are otherwise similar.

The need to learn more about the effects of such mountain obstacles initially led the author to consider what possible propagation paths within the United Kingdom might be studied experimentally to demonstrate the difference between obstructed and unobstructed path propagation. Ideally what was needed was a high-power transmitter so placed that long-distance measurements of its emissions over very smooth terrain, or preferably over sea, could be compared with measurements of the same emissions at a similar distance over an adjacent path obstructed by a single mountain obstacle. The paths initially chosen for tests were in Cardigan Bay and the Irish Sea. Later tests, which form the main subject of the paper, were made over longer paths between England and Northern Ireland, using the Isle of Man as the diffracting ridge.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Jowett is in the Post Office Engineering Department.

(1.2) Outline Description of Experiments

An early series of measurements was made at Holyhead, the field strengths being recorded for a period of nine months of signals from the B.B.C. Band I television transmitters at Douglas, Isle of Man, and at Blaen Plwyf, West Wales. This experiment was of limited value because the path lengths were relatively short and also because of differences in the planes of polarization employed and in the geographical locations of the two paths.

The opportunity to undertake a more refined experiment occurred in connection with the development of the I.T.A. television service. The I.T.A. are assigned the use of Channel 9 at both Winter Hill, Lancashire, and at the station under construction at Black Mountain, Belfast. This co-channel share between the Belfast and Winter Hill stations would provide rather more than the accepted standard protection against mutual interference if the normally used propagation data were applicable. However, the paths between the two service areas are unusual because in some directions almost completely over-sea propagation is involved while, in others, over-sea propagation is interrupted only by the mountain ridge of the Isle of Man. Practical propagation data over these two types of path were needed to confirm that in neither of these typical cases would co-channel interference be caused for more than an acceptably small percentage of the total time.

These and other requirements led to a decision to record the field strength set up by the Winter Hill television sound transmissions at three points in Northern Ireland, commencing in early 1958. Later in that year a special series of measurements was also undertaken, involving the use of a mobile field-strength measuring unit at various points along the east coast of Northern Ireland.

In all, therefore, a total of six series of related measurements have been made, five over fixed paths across the Irish Sea, details of which are given in Table 1, and one using mobile measuring equipment. A map showing the location of each of the paths studied is given in Fig. 1, while Fig. 2 shows the profiles of the five fixed paths based on an effective earth radius equal to four-thirds of the actual radius. In Fig. 3, a projection of the profile of the mountain ridge of the Isle of Man is shown, with an indication of the parts of the ridge traversed by each of the direct paths between Winter Hill and the three fixed Northern Ireland sites.

(1.3) Further Comments on Choice of Receiving Sites

The comparison of the receptions of the Douglas and Blaen Plwyf Band I transmissions carried out between June, 1957, and February, 1958, employed a site at the Post Office v.h.f. radio station at South Stack near Holyhead, Anglesey. This site is on high ground and is not typical of domestic television-receiving sites. This, however, was not of great importance since the particular experiments were primarily designed to obtain more general information on the propagation over mountain ridges and were not immediately applicable to specific broadcasting problems.

Table 1

DETAILS OF PROPAGATION MEASUREMENTS OVER FIXED PATHS

Experiment number	Transmitting station					Receiving terminal		Path length	Height of obstructing ridge	Distance from transmitter to obstructing ridge	Period of observations
	Location	Effective radiated power	Frequency	Aerial height above mean sea level	Plane of polarization	Location	Aerial height above mean sea level				
		kW	Mc/s	ft			ft	km	ft	km	
1	Carnmore Douglas	0·06	63·25	780	Vertical	South Stack, Holyhead	460	95	None		1.6.57–28.2.58
2	Blaen Plwyf	0·25	53·25	985	Horizontal	South Stack	460	112	800	70	1.6.57–28.2.58
3	Winter Hill	25	191·25	1 900	Vertical	Newtownards, N. Ireland	50	237	1 600	148	1.2.58–30.11.58
4	Winter Hill	25	191·25	1 900	Vertical	Ballywalter, N. Ireland	35	212	1 600	148	1.4.58–31.3.59
5	Winter Hill	25	191·25	1 900	Vertical	Dundrum, N. Ireland	55	233	480	156	1.2.58–31.1.59

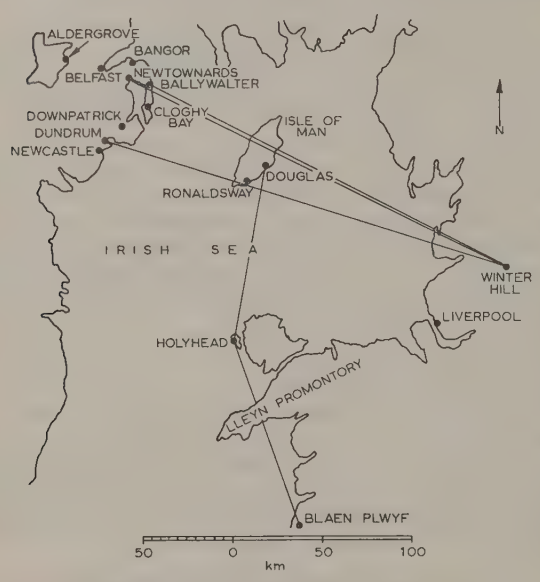


Fig. 1.—Map showing paths studied.

The last three fixed-path experiments were designed particularly with a Band III broadcasting problem in mind. The town of Newtownards was first selected as being a typical town near the east coast of Northern Ireland at which co-channel interference might be observed; its position is also such that the direct path from Winter Hill crosses the Isle of Man ridge at a point about 1 600 ft above sea level. The aerial was erected on the roof of the local Post Office repeater station, but later events proved that the shadowing caused by nearby buildings and slowly rising ground to the east have added a substantial amount of local attenuation to the transmission path from Winter Hill.

Subsequently it was decided to add another receiving site at Ballywalter, within 150 ft of the coast-line, to study a path obstructed by the mountain ridge yet not having the local shadowing effects experienced at Newtownards. Apart from this difference in local effects and a path shortening of some 25 km, the path from Winter Hill to Ballywalter is practically identical with that from Winter Hill to Newtownards.

The site at Dundrum was selected to give as nearly as possible a completely over-sea path of similar length to the other two. Ideally it would have been desirable to select a path just to the south of the Isle of Man, but various local factors favoured the choice of Dundrum, with a path crossing the southern tip of the island where the height is only 470 ft. It is thought that the results obtained are closely similar to those that would have been obtained over an over-sea path completely unobstructed apart from the curvature of the earth.

The site at Ballywalter and, as far as possible, that at Dundrum were deliberately selected as low lying and locally unobstructed towards Winter Hill. From measurements made at these two sites it was thought that the field strengths over a much wider area could be derived by the application of reasonable values of height gain and local shadowing factors. In fact, as already stated, a series of measurements with mobile equipment was also made to assist in this respect.

At all the sites selected a half-wave dipole aerial of appropriate polarization was used 30 ft above local ground, as normally assumed for television broadcast reception.

(2) DATA AVAILABLE FOR PREDICTION OF FIELD STRENGTH SET UP OVER NON-OPTICAL PATHS

(2.1) Smooth-Earth Curves

A number of workers have undertaken the calculation of field strengths to be expected by wave diffraction round a smooth spherical earth.[7–9] Recently an atlas of v.h.f. propagation curves has been published[10] which gives results based on the methods of Van der Pol and Bremmer.[7] In these curves tropospheric effects are allowed for only to the extent of assuming a ray curvature based on standard atmospheric refraction.

The variation of field strength with distance, for a frequency of 150 Mc/s, vertical polarization and over-sea transmission, is shown in curve (*a*) of Fig. 4. A transmitting terminal height of 500 m (compared with the actual height of 580 m) and a receiving terminal height of 10 m have been taken in reading the curves.

(2.2) Empirical Curves

It has been known for a number of years that any theoretical curves of the field strength of v.h.f. and higher-frequency signals based on the diffraction of waves round a smooth earth, even if allowance is made for standard atmospheric refraction, give values which fall far below those measured at distances well beyond the horizon.

The C.C.I.R. IXth Plenary Assembly at Los Angeles, 1959 has recommended the use of empirical curves giving the predicted levels of field strength set up for 1%, 10% and 50% of

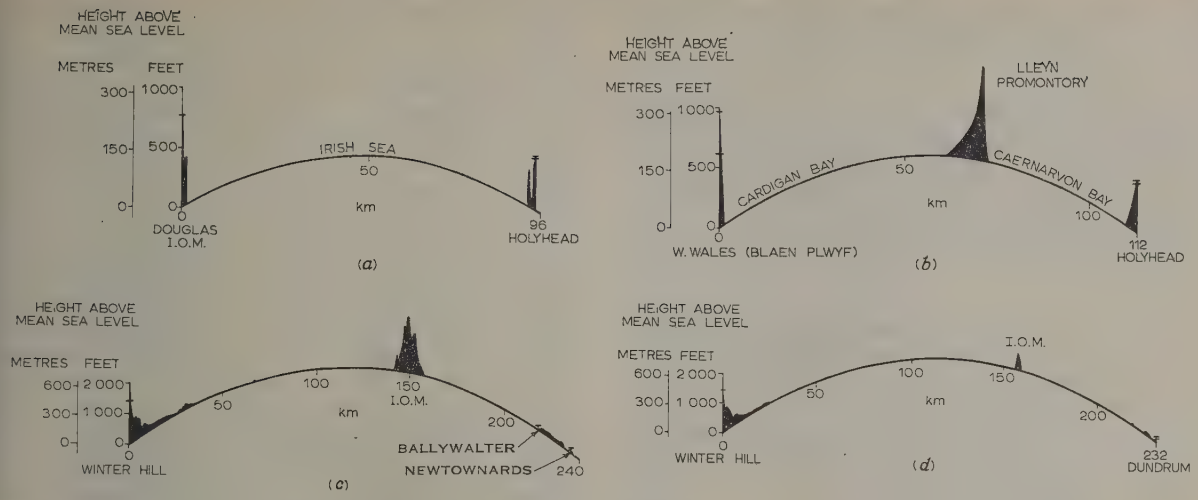


Fig. 2.—Path profiles.

(*a*) Caramore, Douglas, to Holyhead.
(*b*) Blaen Plywf to Holyhead.
(*c*) Winter Hill to Newtownards, also showing position of Ballywalter.
(*d*) Winter Hill to Dundrum.

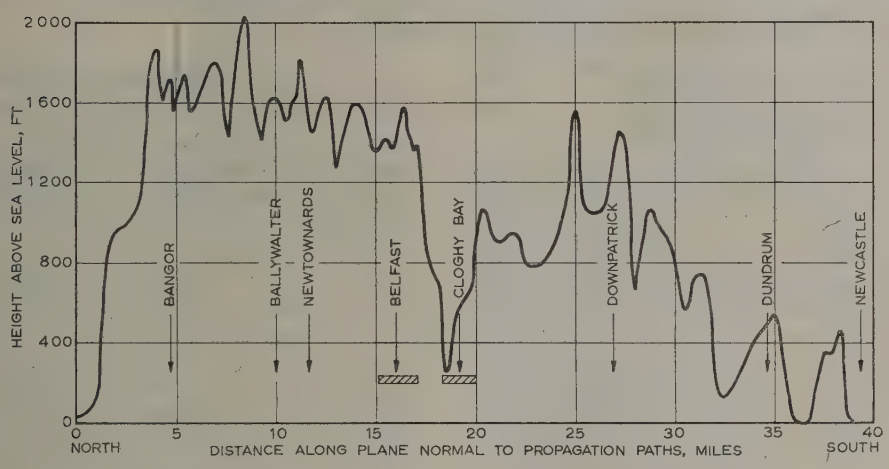


Fig. 3.—Projection of profile of Isle of Man ridge normal to propagation paths.

he time at long distances from the transmitter. These are lesigned to give planning engineers statistical data on the levels f interfering field strengths likely to be set up at long distances o that adequate separations can be arranged between co-channel tations. The C.C.I.R. 50% empirical curve is reproduced as urve (*b*) in Fig. 4 and is markedly different from that predicted rom smooth-earth diffraction theory [curve (*a*)]. For clarity, he 10% and 1% tropospheric curves are not shown, but they nay readily be deduced by adding 9 and 18 dB, respectively, to he 50% curve. It is known that most of the paths which were tudied in order to derive these empirical curves are over-land aths, and there is some evidence to show that the fading range ver long paths is greater when they lie largely or completely ver sea. Hence it need not be expected that the measured esults at present being reported will entirely accord with these mpirical curves.

(2.3) Estimation of Field Strengths likely to be set up over Paths Obstructed by a Mountain Obstacle

Referring to Table 1, paths 2, 3 and 4 lie across what may e regarded as almost a knife-edge mountain ridge near mid-

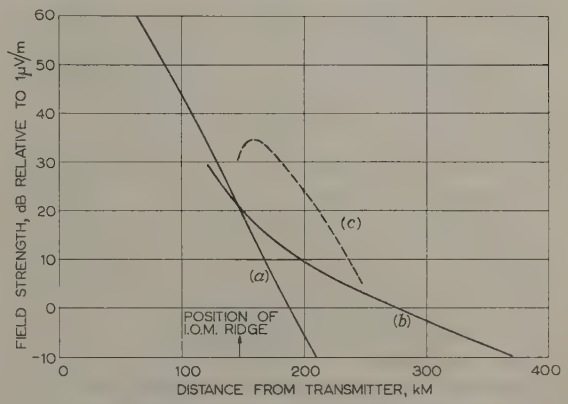


Fig. 4.—V.H.F. propagation curves.

(*a*) C.C.I.R. atlas ground-wave curve, 150 Mc/s, vertical polarization.
(*b*) C.C.I.R. empirical tropospheric curve (Los Angeles, 1959) showing field strengths exceed for 50% of the time, adjusted for 1900ft transmitting aerial heigh
(*c*) Derived curve, based on simple knife-edge diffraction theory.

path. A number of observations on similar paths have been reported from various parts of the world and lead to the following general conclusions:

(i) In favourable circumstances the average field strength set up may exceed not only that predicted by smooth-earth diffraction theory but also that given by empirical curves for propagation over terrain having only small or moderate irregularities.
(ii) The range of signal fading experienced on such paths is generally appreciably less than that for similar paths having no mountain obstacle.
(iii) The increased signal strength and reduced fading range are particularly pronounced in the case of transmissions in the lower part of the v.h.f. range.

It is believed that the field strength set up over such paths may be critically dependent upon the overall path profile, the effects mentioned above being greatest when, as for paths 2, 3 and 4, the mountain ridge is relatively sharp and extends for no great distance along the direction of propagation. It is also probable that the so-called obstacle-gain effects are greatest when both transmitting and receiving terminals are relatively low, because in this case the transmission loss without the intervening mountain ridge would usually be very large.

A simplified approach to the problem is to treat the overall transmission loss as compounded of four parts, namely:

(*a*) The free-space transmission loss over the total distance.
(*b*) The loss caused by diffraction around the ridge itself, treated as a knife-edge.
(*c*) The loss over and above free-space loss existing on a hypothetical link between the transmitter and the ridge.
(*d*) The loss over and above free-space loss over a hypothetical link between the ridge and the receiving terminal.

In the case of the ridge-obstructed path 2 it can easily be shown that both the excess losses (*c*) and (*d*) are very small, if not negative, since the clearances above the sea on both sides of the ridge are such that, allowing for the 180° phase change at reflection, the sea-reflected ray and the direct ray are practically in quadrature. The same result applies under standard refraction conditions for the Winter Hill—Isle of Man section [excess loss (*c*)] of paths 3 and 4. However, the final sections [excess loss (*d*)] of these two paths, i.e. the sections from the Isle of Man ridge to Newtownards and Ballywalter in Northern Ireland, give considerably less than first-Fresnel-zone clearance, and the path losses there may be expected to exceed their corresponding free-space losses by substantial amounts which, moreover, will vary with the varying refractive-index gradient in the lower troposphere.

In all cases the free-space loss (*a*) is, of course, readily calculable while the diffraction loss (*b*) may be calculated using the parameter $2\pi h/\sqrt{[\lambda d_1 d_2/(d_1 + d_2)]}$ which enters into the diffraction formula. In this parameter, which applies specifically to the case of knife-edge diffraction, d_1 and d_2 are the distances from the transmitting and receiving terminals to the ridge; $d_1 + d_2 = d$, the total distance; and h is the effective ridge height. These two losses have been calculated for the case of signals from Winter Hill passing over the Isle of Man ridge at a height of 1 600 ft (i.e. 500 m) and being diffracted to receiving points of effective heights 10 m at a range of distances beyond the ridge.

The excess loss over free space for paths 3 and 4 is, as already stated, around zero and possibly negative, because the clearance over the Winter Hill—Isle of Man sections is sufficient to give nearly a −90° phase difference between direct and sea-reflected rays at the 1 600 ft ridge. In fact, for standard refraction conditions and perfect reflection, the loss is estimated to be −2·5 dB, but some separate measurements of Winter Hill reception at a high site in the Isle of Man have suggested that the conditions of standard refraction are not met and it appears more realistic to assume that loss (*c*) is virtually zero for both paths. Thus approximately free-space field conditions are assumed to be set up just above the 1 600 ft ridge of the Isle of Man.

It now remains to allow for the excess loss (*d*) because of the effects of sea reflections on the paths of both parts to the west of the Isle of Man ridge. This has been done by reference to the curves given in the C.C.I.R. atlas of ground-wave curves and assuming that the ridge acts as a transmitter having an effective height above ground of 500 m. The curves have been used by reading off, for a range of distances, the difference between the free-space and the predicted ground-wave field, and the additional loss so derived has been added to the overall free-space loss (*a*) and ridge-diffraction loss (*b*) already calculated. This has enabled a curve of field strength versus distance to be drawn, shown as curve (*c*) of Fig. 4, from which the well-known recovery effect appearing just after the ridge will be clearly seen.

This prediction method has been considered in some detail because it suggests a means by which the field strengths set up over paths obstructed by a mountain ridge can be approximately calculated. Certain changes in the detailed calculations would be necessary in particular cases; for example, it would not be appropriate to use the C.C.I.R. atlas where the terminal points are not visible from the ridge or where the terrain is not substantially smooth. But the agreement of this method with results, quoted in a later Section, gives some confidence in this general approach.

(3) RESULTS OF TESTS

(3.1) Field Strengths Measured at Holyhead (Paths 1 and 2)

Using the results obtained at Holyhead on signals from Blaen Plwyf and Douglas, the median signal levels for each hour of the day for each of the nine months of observations were determined. The distribution of these hourly median signal levels, taking each month separately, was then plotted and the results for the unobstructed path 1 (Douglas to Holyhead) and the obstructed path 2 (Blaen Plwyf to Holyhead) are shown in Fig. 5. In both cases the results have been normalized to show the field strength set up by a radiated power of 1 kW. A comparison of these two presentations shows very clearly the effect of the mountain ridge in path 2 in reducing the fading range, both within any one month and from month to month, although the difference in planes of polarization employed must not be disregarded. However, the evidence available suggests that this would have had no more than a secondary influence on the comparative results and that diffraction over the Lleyn promontory in the case of path 2 is the main cause of the large difference in the fading characteristics over the two paths.

The distributions of hourly median levels, taking the nine-month period as a whole, are shown in Fig. 6. From this it will be seen that the overall median signal level recorded over the obstructed path from Blaen Plwyf exceeds that over the unobstructed Douglas path by about 3 dB. For small percentages of the time, however, the Douglas path gave slightly larger signals than did the Blaen Plwyf path. It is of interest to note that the median signal level over the Blaen Plwyf path was some 16 dB below the level of free-space signal, while that for the Douglas path was 19 dB below its corresponding free-space signal. Although local siting effects at Holyhead may have contributed in part to these losses (more particularly for the Blaen Plwyf path for which the foreground at Holyhead did not fall away sharply) it may be taken that these losses are largely those due to the large-scale geometry of the two paths. In particular, the knife-edge diffraction formula already quoted gives a loss over the ridge of 16·5 dB, and it might be expected that the overall loss relative to free space would also approximate to this on path 2 for the reasons stated in Section 2.3. Thus

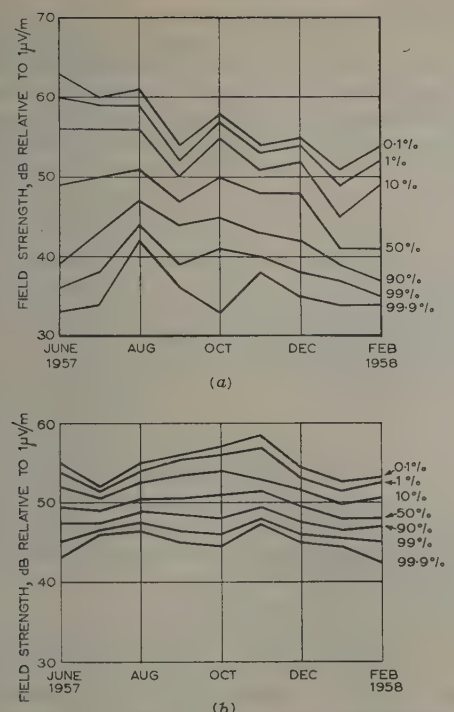


Fig. 5.—Distribution of hourly median values of field strength.
(*a*) Douglas to Holyhead.
(*b*) Blaen Plywf to Holyhead.

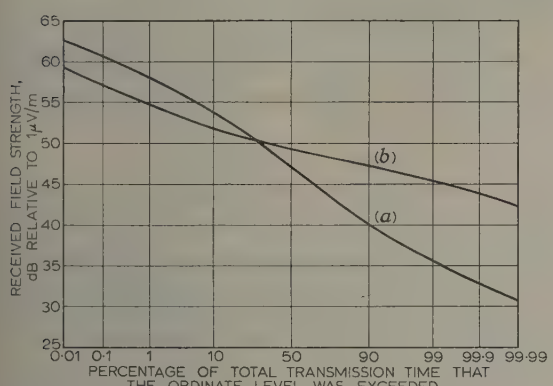


Fig. 6.—Overall distribution of hourly median field strengths recorded over period June, 1957, to February, 1958.
(*a*) Douglas to Holyhead.
(*b*) Blaen Plywf to Holyhead.

the results for this path conform with expectations based on this simple knife-edge theory.

When signal strengths were generally low, no doubt under conditions of sub-refraction, the difference between the two paths was very marked. Thus it will be seen that the signal strength set up over the unobstructed path fell below that for the obstructed path by 10 dB for the worst 1% of the total time. In fact, the fading range between the 0·1% and 99·9% points of the distribution curves was 28 dB for the completely over-sea path from Douglas and only 13 dB for the obstructed Blaen Plwyf path.

(3.2) Field-Strength Measurements of Winter Hill at Fixed Sites in Northern Ireland (Paths 3, 4 and 5)

(3.2.1) Monthly Analyses of Quarter-Hourly Median Values.

Chart recordings have been made over most of the period February, 1958, to January, 1959, at the Newtownards, Ballywalter and Dundrum sites to show the field strengths set up by the Winter Hill sound transmissions. The normal chart speed employed was 1 in per hour, which has permitted the median signal level over approximately each 15 min period to be read off. For each path, taking one month at a time, the distributions of these values have been deduced and are shown* plotted against the month of observation in Fig. 7. Since interest centres

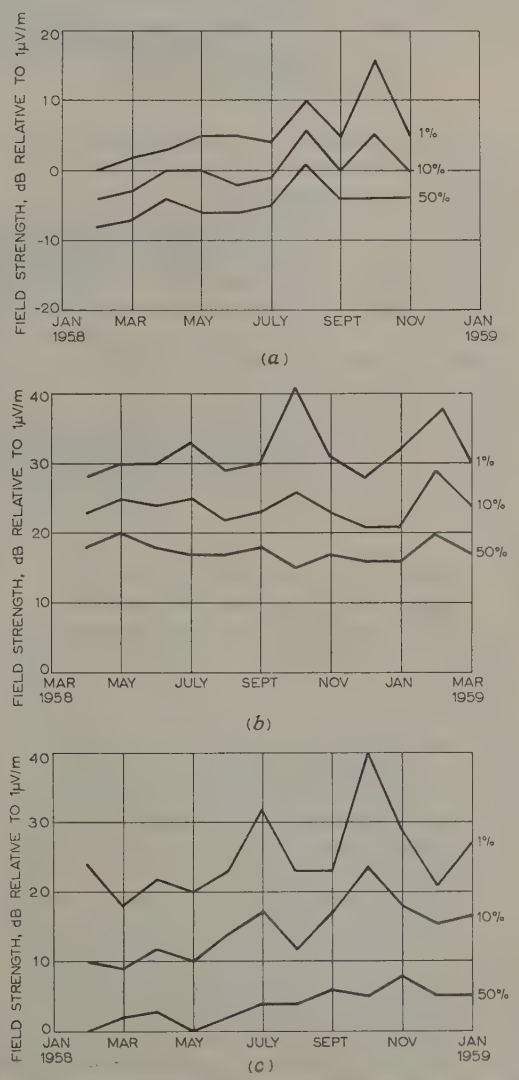


Fig. 7.—Distribution of 15 min median values of field strength.
(*a*) Winter Hill to Newtownards.
(*b*) Winter Hill to Ballywalter.
(*c*) Winter Hill to Dundrum.

mainly on the occasionally occurring high levels and on the median overall levels, these monthly distributions have not been carried down to give field strengths exceeded for 90% and 99%

* All field strengths shown in the Figures have been normalized to correspond to the condition of 1 kW (e.r.p.).

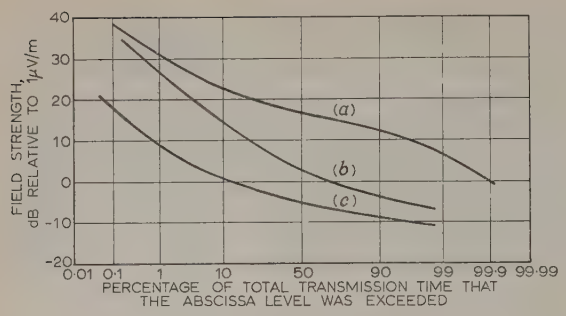


Fig. 8.—Overall distributions of hourly median field strengths.
(*a*) Winter Hill to Ballywalter.
(*b*) Winter Hill to Dundrum.
(*c*) Winter Hill to Newtownards.

of the total time. However, the overall distributions for the total periods of observations which are shown in Fig. 8 have been extended to cover these levels.

From Fig. 7 it will be seen that the almost unobstructed over-sea path to Dundrum shows a considerably greater degree of fading than do the other two paths obstructed by the mountain ridge. Furthermore, the overall median level of signal recorded at Dundrum is 7 dB above that at Newtownards and 14 dB below that at Ballywalter. In making a comparison of path losses, however, the local site conditions as well as the differences in path lengths have to be considered. Thus the 21 dB difference between the Ballywalter and Newtownards median levels is undoubtedly largely the result of local shielding, partly by nearby buildings and certainly by rising foreground at the Newtownards site; it is also partly due to the difference in distance of 25 km. Reference to the empirical C.C.I.R. data of median field strength in Fig. 4, curve (*b*), shows that the median field of +2·5 dB relative to 1 μV/m measured at Dundrum is very close to the predicted value of +5 dB relative to 1 μV/m at 233 km. The measured median field strength at Ballywalter is, however, 11·5 dB above that indicated by the C.C.I.R. curve.

The difference between the median values of field strength set up at Dundrum and at Ballywalter merits closer attention. The heights and immediate local foregrounds of the two sites are closely similar, but the effect of a small land promontory rising to a height of 100–150 ft one or two miles to the East of Dundrum appears to have offset its otherwise good situation for the reception of Winter Hill signals. In fact, as explained in Section 3.2.3, the site at Dundrum is little better than average for the surrounding area, and signals some 12–15 dB higher were measured on two particular days at a very similar site open to the sea to the east of the land promontory. In other words, it appears that had the Dundrum site been completely open to the sea it would have recorded median levels of signal practically as high as those recorded at Ballywalter, i.e. these levels would have been about 10–12 dB above those indicated by the empirical C.C.I.R. median curves. Nevertheless, this apparent equality of median transmission loss that would have been observed over the two paths is probably fortuitous, since there is ample evidence to suggest that the predominant transmission modes over the two paths are quite different.

For example, the considerably shallower distribution curve of the Ballywalter signals as compared with that of the Dundrum signals clearly suggests that the signals received at Ballywalter are caused mainly by diffraction over the mountain ridge whereas those at Dundrum are much more subject to changing tropospheric conditions. Thus the Dundrum signal levels observed for 1% of the time reach to within 6 dB of the Ballywalter 1% signals; on the other hand the Dundrum 99% signals fall below the Ballywalter 99% signals by nearly 20 dB. That the field at Ballywalter corresponds to that set up by ridge diffraction receives confirmation from the study of fading characteristics discussed in Section 3.2.2. It is also interesting to see from Fig. 4 that the calculated ridge-diffracted field at this distance is only 1·5 dB above the median value actually observed at Ballywalter.

The similarity of shape of the distribution curves for Ballywalter and Newtownards is also of some interest. Both are affected by the mountain ridge and show a flattening out between the 90% and 10% levels. The range of fading between these levels is, in fact, only about 10 dB as compared with 17 dB for the Dundrum path. In both the Ballywalter and Newtownards cases, however, the distribution curves steepen for the occasionally observed higher levels, and this is thought to be due to tropospheric reflections which might occasionally predominate over the ridge-diffracted signal. Again, for the Ballywalter path, the curve also appears to steepen for the lowest levels. This may well be due to the rapidly increasing losses that may be expected to occur when sub-refraction reduces appreciably the path clearances on either side of the mountain ridge. It would not have been possible to observe any similar effect on the Newtownards path because of the limit placed by receiver sensitivity at the lowest recorded levels.

When the month-to-month variations shown in Fig. 7 are examined it is clear that the median observed levels on all paths did not vary greatly. The median levels for the Ballywalter path varied, in fact, by no more than 2 dB. On the other hand, the field strengths exceeded for 1% of the time showed marked changes especially on the over-sea path to Dundrum. Here a maximum of 40 dB above 1 μV/m [for 1 kW (e.r.p.)] was recorded in October, 1958, and a minimum of +18 dB for March of that year, this seasonal variation according with other reported observations over long v.h.f. paths. The Ballywalter 1% signals showed a minimum level of +27 dB in December and a maximum level of +41 dB in October, while the seasonal variations observed at Newtownards fell between these two extremes.

(3.2.2) Some Characteristics of the Fading Observed.

The general statistics of the field strengths measured and their distributions have already been discussed in Section 3.2.1. Certain characteristics of the signal recordings are, however, of particular interest. Those obtained over the Dundrum path normally showed the expected type of fairly fast fading, the fading rate being generally greatest when the signal strength was lowest, and the short-term fading range often extending over a range of from ±5 to ±10 dB.

The recordings obtained over the ridge paths to Ballywalter and Newtownards were of a different character. On some days the short-term fading was pronounced and rapid, but more usually the range of fading was not greater than ±5 dB and frequently was less than ±2 dB, especially in the case of the Ballywalter path. On numerous occasions the fading range would decrease to a very small value, and a very thin trace, showing practically no fading, would persist for one or more hours. Under these conditions it has been assumed that the arriving signal was almost entirely that due to diffraction by the mountain ridge and that the contribution from partially reflecting atmospheric layers was negligible. This is borne out by the results of a rough analysis of such occasions observed during December, 1958, at Ballywalter. Thus the total period when the signal trace thinned to show less than ±1 dB variations was 26 hours. The actual signal level on all these occasions was always within 2 dB of +16 dB relative to 1 μV/m [for 1 kW

(e.r.p.)]. This figure, it may be noted, is also the median value of field strength observed throughout the month of December.

Further evidence in support of this is given by the conditions simultaneously observed on the almost unobstructed path to Dundrum. It has been noticed that pronounced thinning of the trace on the Ballywalter recording is frequently accompanied by a definite reduction in the field strength recorded at Dundrum. A very clear example of this is given in Fig. 9. Here both the than the median value of those recorded in an area of some 10 miles radius centred on Ballywater.

The Dundrum site was found to be more or less representative of nearby inland and coastal sites in the southern area, which was little affected by the Isle of Man ridge.

The scatter of measurements is considerable, largely owing to the hilly nature of some parts of the terrain. A particular observation was that points along the Ards peninsula which

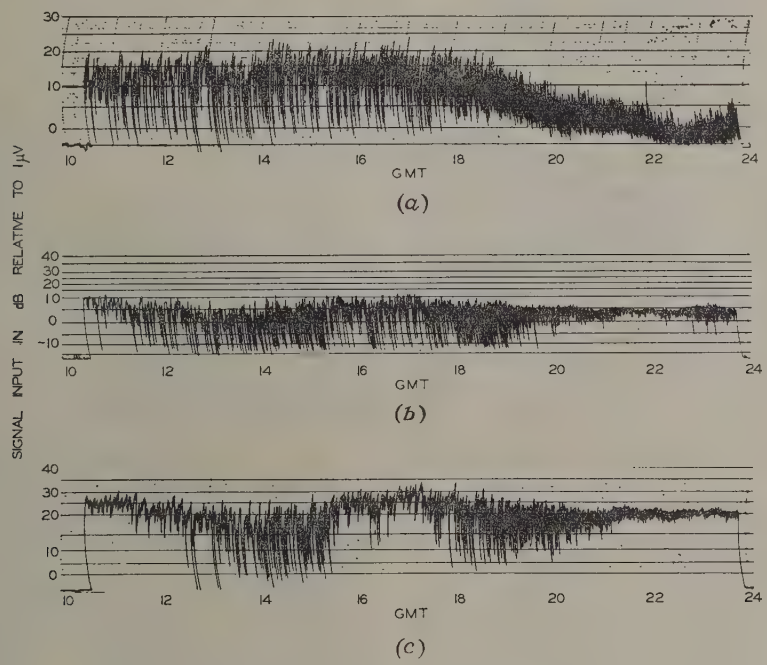


Fig. 9.—Recordings of Winter Hill transmissions, 12th October, 1958.

(*a*) Reception at Dundrum.
(*b*) Reception at Newtownards.
(*c*) Reception at Ballywalter.

Ballywalter and Newtownards recordings exhibit trace thinning at about the same time, while the Dundrum signal falls by about 10 dB for nearly the same period. The evident conclusion is that the components of the received signal set up by tropospheric reflections fell to low levels nearly simultaneously on all paths, giving low fields at Dundrum but leaving the ridge-diffracted field at Ballywalter and Newtownards.

(3.2.3) Mobile Field-Strength Measurements in Northern Ireland.

Measurements using mobile receiving equipment were made primarily to determine what relation the measurements at Newtownards, Ballywalter and Dundrum might bear to the field strengths in other parts of Northern Ireland to the east of Belfast. By making measurements at a large number of places and relating them to measurements recorded simultaneously at the Ballywalter and Dundrum sites it is possible to estimate, over a wide area, the likely effects of any potential co-channel interference from Winter Hill.

Ballywalter, as had been expected, was found to be a better than average site, and it was noted that the measured field strength fell off noticeably as the vehicle proceeded inland until much higher ground was reached at Belfast. The field strengths recorded at Ballywalter, in fact, proved to be about 6 dB greater were obstructed by very high ground at the Isle of Man generally showed higher field strengths than did other areas such as Cloghy Bay, Dundrum, where only low obstructions to a height of about 400 ft applied.

It may be concluded that, at sites to the east of Belfast and along the Ards peninsula, field strengths centred on levels about 6 dB lower than those measured at Ballywalter will be set up. The effect of undulating terrain will be to give a 'spread' of ± 10 dB to these general results. At coastal and near-coastal sites, which are almost unobstructed by the Isle of Man ridge, the field strengths set up will be generally of the same order as those measured at Dundrum, though again a ± 10 dB spread of results is to be expected according to the disposition of the receiving site relative to local terrain. On this basis Table 2 has been prepared to indicate the expected levels of interfering signal from a 100 kW transmission from Winter Hill (corresponding to the e.r.p. of the vision signals) at points on and near the east coast of Northern Ireland.

(3.2.4) Extent of Protection of Black Mountain Viewers from Winter Hill Interference.

If it is first assumed that a 30 dB protection ratio must be given at the input of the receiver to a fringe-area wanted field of

Table 2

INTERFERING STRENGTHS LIKELY TO BE EXCEEDED AT SITES IN NORTHERN IRELAND FOR STATED PERCENTAGES OF TIME

	Field strength*								
	At poor receiving sites			At typical receiving sites			At good receiving sites		
	1% of time	10% of time	50% of time	1% of time	10% of time	50% of time	1% of time	10% of time	50% of time
	dB	dB	dB	dB	dB	dB	dB	dB	dB
Paths obstructed by high I.o.M. ridge	35	27	21	45	37	31	55	47	41
Paths almost unobstructed by I.o.M. ridge	31	18	7	41	28	17	51	38	27

* In dB relative to 1 μV/m for a 100 kW transmission.

about 500 μV/m (54 dB relative to 1 μV/m), the interfering field strength should not exceed +24 dB relative to 1 μV/m for more than a small percentage of the time, given no aerial discrimination against the unwanted signal. However, some appreciable aerial discrimination can be safely assumed because of (*a*) the directive properties of typical receiving aerials and (*b*) the fact that Winter Hill employs vertical polarization whereas Black Mountain signals will be horizontally polarized. Some recent unpublished measurements suggest that a safe planning figure to accept for the resultant of these two effects is of the order of 18 dB. This means, in effect, that the interfering field should be at least 12 dB below the limiting wanted field of +54 dB relative to 1 μV/m.

Taking first the case of typical receiving sites we see from Table 3 that the 1% interfering fields are estimated to be +45 and +41 dB relative to 1 μV/m, the former being the figure for paths obstructed by the Isle of Man ridge. It may thus be concluded that for such obstructed paths, average sites on a 500 μV/m median wanted-field-strength centred near the east coast of Northern Ireland may experience interference from Winter Hill for about 2–3% of the total time; for unobstructed paths, average sites may experience interference for rather less than 1% of the total time. These figures, which assume that reasonable precautions are taken in respect of the aerial installation, are reassuring in that such rates of occurrence of noticeable interference in fringe areas are not worse than those generally accepted in national and international television planning.

Note must, of course, be taken of the fact that sites may experience interference levels about 10 dB higher than these. Such sites, however, will often be placed so as to receive a significantly higher than average wanted signal strength, and this will largely compensate for the higher levels of interfering signals. In addition, it is expected that the Black Mountain transmitter will establish field strengths along the east coast of Northern Ireland which are on average considerably greater than the limiting 500 μV/m. The overall conclusion is therefore that the extent of interference by Winter Hill to the Black Mountain service area is likely to be small and below the generally accepted limit of co-channel interference. Such slight interference as does occur may be rather greater for sites which are affected by Winter Hill transmissions by propagation over the Isle of Man ridge.

(4) COMPARISONS OF RADIO FIELD-STRENGTH MEASUREMENTS WITH METEOROLOGICAL DATA

(4.1) Data Studied

An attempt has been made to relate the values of field strength measured over paths 3, 4 and 5 from Winter Hill to Northern Ireland to meteorological observations along this path. In the first place, data giving the temperature, pressure and dew-point from observations at Liverpool and Aldergrove were used, but later attention was given to these factors measured at Ronaldsway in the Isle of Man. Both aerological and ground measurements were studied, but it was found that the derived refractive-index gradients in the lower atmosphere bore no more marked relationship to the radio-signal levels than did the derived refractive indices at the ground. In consequence, most of the comparisons used only the ground-based meteorological data.

(4.2) Results of Comparisons

Practically all the comparisons were carried out by reference to the radio field strengths observed over path 5 at Dundrum since it was this path which showed the greatest signal-level variations and which was the most clearly subject to changing atmospheric conditions. The results of the comparisons showed that a positive and sometimes quite marked correlation existed between the three quantities, radio field strength at Dundrum, observed atmospheric pressure at the ground and derived refractive index at the ground. Nevertheless, when the data were compared over relatively short intervals of time, e.g. in terms of the three-hourly or daily median values, it was clear that correlation was by no means complete and that factors other than those observed at Ronaldsway must have influenced the attenuation of the radio waves. As has been found by other experimenters, the correlation between these quantities proved to be rather closer when smoothed running-means, taken over a period of several days or a month, were compared.

The results of these analyses thus tend to confirm the views now held by many planning engineers that a close and reliable connection between radio and meteorological data is not to be expected or sought for when considering the transmission of signals over tropospheric paths. This emphasizes the need to collect further radio data over paths covering a representative variety of tropographical and climatic conditions for direct application to further planning work in the broadcasting and communication fields.

(5) CONCLUSION

A description has been given of v.h.f. field-strength measurements made over a number of paths across the Irish Sea. The results have indicated the extent to which the transmission of v.h.f. waves may be affected by mountain obstacles. A particular application of the results has been made to deduce the possible extent of co-channel interference that may exist between the I.T.A. transmissions at Winter Hill and the I.T.A. service to be set up at Black Mountain, Belfast.

(6) ACKNOWLEDGMENTS

A number of colleagues have assisted in setting up the apparatus for these measurements and in analysing the results. The author wishes to thank them for the very considerable help that has thus been given. Acknowledgment is due to the Engineer-in-Chief of the Post Office for permission to publish the paper.

(7) REFERENCES

(1) BULLINGTON, K.: 'Radio Propagation at Frequencies above 30 Mc/s', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 132.

(2) SCHELLENG, J. C., BURROWS, C. R., and FERRELL, E. B.: 'Ultra Short Wave Propagation', *ibid.*, 1933, **21**, p. 427.

(3) DICKSON, F. H., EGLI, J. J., HERBSTREIT, J. W., and WICKIZER, G. S.: 'Large Reductions of V.H.F. Transmission Loss and Fading by the Presence of a Mountain Obstacle in Beyond Line-of-Sight Paths', *ibid.*, 1953, **41**, p. 967.

(4) KONO, T., HIRAI, M., and UYESUGI, Y.: 'Uses of Diffraction Effect of Mountains for VHF Radio Communications', *Journal of the Radio Research Laboratories*, Japan, 1955, **2**, No. 9, p. 293.

(5) KIRBY, R. S., DOUGHERTY, H. T., and MCQUATE, P. L.: 'Obstacle Gain Measurements over Pikes Peak at 60 and 1 046 Mc/s', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1467.

(6) CRYSDALE, J. H., *et al.*: 'An Experimental Investigation of the Diffraction of Electromagnetic Waves by a Dominating Ridge', *Transactions of the Institute of Radio Engineers*, 1957, **AP5**, p. 203.

(7) VAN DER POL, B., and BREMMER, H.: *Philosophical Magazine*, 1937, **24**, pp. 141 and 845; 1938, **25**, p. 817; 1939, **27**, p. 161.

(8) ECKERSLEY, T. L., and MILLINGTON, G.: *Philosophical Transactions of the Royal Society*, 1938, **237**, p. 273.

(9) NORTON, K. A.: *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 1367; **25**, p. 1203: 1941, **29**, p. 623.

(10) 'Atlas of Ground-Wave Propagation Curves for Frequencies between 30 and 300 Mc/s' (International Telecommunications Union, Geneva, 1955).

621.317.34.029.6

The Institution of Electrical Engineers
Paper No. 3155 E
Mar. 1960
©

A METHOD OF AMPLITUDE AND PHASE MEASUREMENT IN THE V.H.F.–U.H.F. BAND

By G. D. MONTEATH, B.Sc., A.Inst.P., D. J. WHYTHE, B.Sc.(Eng.), Associate Members, and K. W. T. HUGHES, B.Sc.

(*The paper was first received* 24*th March, and in revised form* 21*st August*, 1959.)

SUMMARY

A null method has been devised for measuring changes in the amplitude and phase of the transmission characteristic of a network at any frequency in the range 41–1 000 Mc/s. A commercial instrument, designed for admittance measurement, is used with only slight modification for this application. Its subsequent use as an admittance meter is unaffected.

The measurement of phase can be made to within about ±3°, unless a large variation of amplitude is encountered, when reading accuracy limits the accuracy of phase measurement in the regions of small amplitude. If the error in a measurement is regarded as a vector, then, with any amplitude variation, the maximum value of the magnitude of this error vector is about 6% of the full-scale reading of the instrument.

(1) INTRODUCTION

The paper describes a null method of measuring changes in the amplitude and phase of the transmission of a 2-terminal-pair network at any frequency between 41 and 1 000 Mc/s. The method was devised for measuring the radiation patterns of model aerials, and it is particularly suitable for use when the insertion loss of the network is high.

Three complex quantities are necessary to specify a linear and reciprocal 2-terminal-pair network (these may, for example, be the self-impedance at each terminal pair and the mutual impedance between terminal pairs) and, in general, a change in the network will cause all three quantities to change. But if the impedance at each terminal pair is constant and independent of the impedance connected to the other terminal pair (this condition implies a high attenuation in the network), it is sufficient to measure changes in the transmission of the network in terms of one complex quantity. (For example, it is immaterial whether mutual impedance or mutual admittance is measured, since these quantities will vary in a constant ratio.) Moreover, this quantity will be independent of the impedances in which the network is terminated, apart from a constant factor. The method of measurement to be described is primarily applicable to networks satisfying these conditions.

Several methods have been devised for measuring the phase difference between two alternating voltages at high frequencies. In one system of use in the v.h.f. and u.h.f. bands, the output voltage from the network under test is balanced against a subsidiary voltage obtained from the probe of a well-matched slotted line. This subsidiary voltage is adjusted in amplitude by an attenuator and in phase by traversing the probe along the slotted line. A phase meter for microwave frequencies[1] employs a balanced modulator in conjunction with a calibrated phase-shifter in a waveguide section. In a third method,[2] which has been used to compare the outputs of directional couplers, two signals are compared at an intermediate frequency. The output of a local oscillator is fed with an adjustable phase difference to two frequency changers. This phase difference is adjusted to give intermediate-frequency outputs which are in anti-phase, and their amplitudes are compared with a potentiometer. Each of these methods entails the use of a slotted transmission line or some other kind of phase-shifter having a calibration dependent upon frequency.

More recently a commercial instrument[3] has been developed for determining all the linear parameters of a 2-terminal-pair network. The results are obtained in absolute rather than relative form. The method to be described is believed to be more convenient and rapid than other methods in common use. It makes use of a simple instrument designed for admittance measurement, with only a minor modification which does not affect its normal use.

(2) METHOD OF MEASUREMENT

The output from an oscillator is fed to a receiver by two parallel paths; one path includes the network under test and the other includes an instrument whose output can be adjusted in amplitude and phase to cancel the output from the network. This arrangement is shown in Fig. 1.

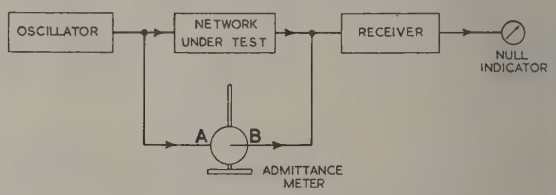


Fig. 1.—Principle of measurement.

A and B (also marked on Fig. 2) are the 'Generator' and 'Detector' terminals of the admittance meter.

The instrument used for controlling the amplitude and phase of the cancelling voltage is a commercial admittance meter,[4] a schematic of which is shown in Fig. 2. Three coaxial transmission lines have a common junction, which is connected to the generator when the instrument is used for admittance measurement. One of these lines is terminated by the unknown admittance; the two others are terminated respectively by a standard conductance and a standard susceptance. Both standards are equal in magnitude to the characteristic admittance of the lines, the standard susceptance consisting either of a $\lambda/8$ or a $3\lambda/8$ short-circuited line or the equivalent in lumped capacitance. The current entering each line is sampled by a loop, which can be rotated in order to vary the amplitude (but not the phase) of the e.m.f. induced in it. The outputs of the three loops are combined and fed to a detector. By rotating the loops, the e.m.f. induced by the current flowing to the unknown admittance can be cancelled by the in-phase and quadrature e.m.f.'s induced in the two other loops. Scales associated with the 'standard' loops enable the unknown admittance to be read

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Monteath and Mr. Whythe are, and Mr. Hughes was formerly, with the British Broadcasting Corporation. Mr. Hughes is now with Racal Engineering, Ltd.

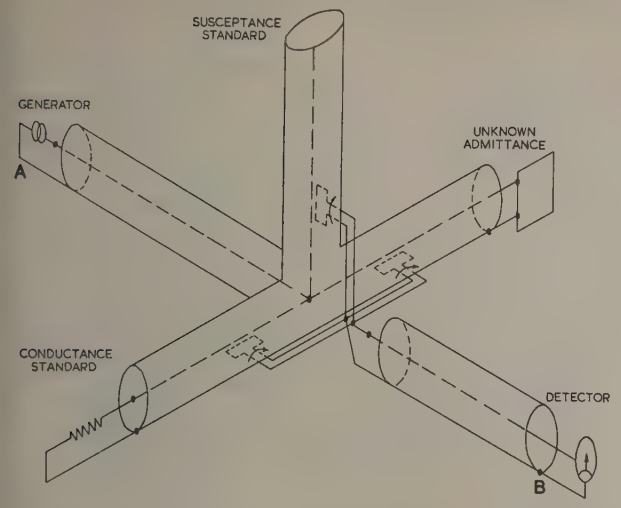


Fig. 2.—General arrangement of the admittance meter as used for admittance measurement.

directly in millimhos; the scale associated with the 'unknown' loop is calibrated in terms of a factor by which the two other scale readings are multiplied.

If the scale-multiplying factor is set to infinity, so that there is no coupling between the 'unknown' transmission line and its associated loop, the output at the detector terminals consists of two components in quadrature whose amplitudes are separately controlled by the settings of the other two loops, and which are proportional to the readings of the conductance and susceptance scales. The instrument is used in this way as a phase meter. Connected as shown in Fig. 1, the output from the detector terminals is arranged to cancel that from the network under test; the in-phase and quadrature components of this output are then proportional to the readings of the conductance and susceptance scales.

The loop coupled to the standard susceptance can be rotated through 180°, but that coupled to the standard conductance can be rotated through only 90°. The use of the conductance and susceptance scales alone can therefore accommodate a phase variation of only 180°. In order to accommodate all phases, a second standard conductance may be connected to the 'unknown' terminals. The loop coupled to this gives an output in anti-phase to that obtained from the conductance loop. The use of a second standard conductance, although convenient, is not essential, since a single standard conductance can be moved from one position to the other according to the sign of the real part required. The unused loop is set to zero coupling (zero millimhos or infinity scale factor).

As supplied by the makers, the admittance meter is designed for use with 50-ohm transmission line, and the values of both the standard conductance and the standard susceptance are therefore 20 millimhos. If the conductance, susceptance and multiplier scales respectively indicate g millimhos, b millimhos and m, and the 'unknown' arm is terminated in a standard conductance of 20 millimhos, the output is equal to $K(g - 20/m + jb)$ where K is a complex constant. It is convenient to superimpose on the multiplier scale a detachable scale with engravings similar to those of the conductance scale. If this reads g' the output is proportional to $K(g - g' + jb)$. The complex constant K remains unknown, in general, throughout the measurements.

(3) EXPERIMENTAL RESULTS

Experiments were performed at 500 and 1 000 Mc/s to assess the accuracy of measurement. The method described in Section 2 was adopted to measure the relative amplitude and phase of the output voltage from the probe of a slotted line for various standing-wave ratios on the line. The results are compared with curves which were calculated from the standing-wave ratios, measured independently against a piston attenuator. At each frequency four terminations were used to give voltage standing-wave ratios between 0·98 and 0·007. A preliminary examination of the standing-wave pattern showed the effect of loss and discontinuities in the slotted line to be negligible.

It is shown in the Appendix that, in the absence of loss, the locus (on the Argand diagram) of the voltage on a uniform transmission line is an ellipse. The way in which position on the ellipse varies with position on the transmission line is illustrated in Fig. 3. A point P on a circle is defined by the

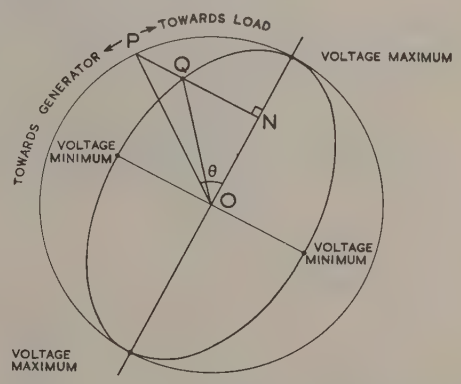


Fig. 3.—Locus of the voltage vector, OQ, on a uniform transmission line.

angle θ, which is equal to the angular distance along the transmission line from the voltage maximum. The point Q on the perpendicular PN to the axis is defined by $QN/PN = s$, where s is the voltage standing-wave ratio. If the probe is moved at a uniform rate along the slotted line, P moves at a uniform rate round the circle and Q traces out an ellipse. The circle is the auxiliary circle of the ellipse.

Measurements are made with greatest accuracy when the whole of the conductance and susceptance scales are used. This occurs when the axes of the ellipse are inclined at approximately 45° on the Argand diagram. To achieve this result, a phase-shifter (a variable length of coaxial line) was placed in series with the admittance meter and adjusted before each measurement commenced. The results of the measurements are compared in Figs. 4 and 5 with the corresponding theoretical points lying on the ellipse. For clarity, overlapping points measured with the probe in different positions on the slotted line have been differently coded. The eccentricity of each theoretical ellipse was determined from the measured standing-wave ratio on the slotted line, while the positions of the points were calculated from the measured displacement of the probe from the voltage antinode. The size and inclination of the theoretical ellipse was adjusted arbitrarily to give the best fit to the experimental points.

(4) MEASUREMENT ERRORS

The source of greatest error in phase was found to be that of scale reading, particularly when the network attenuation varied

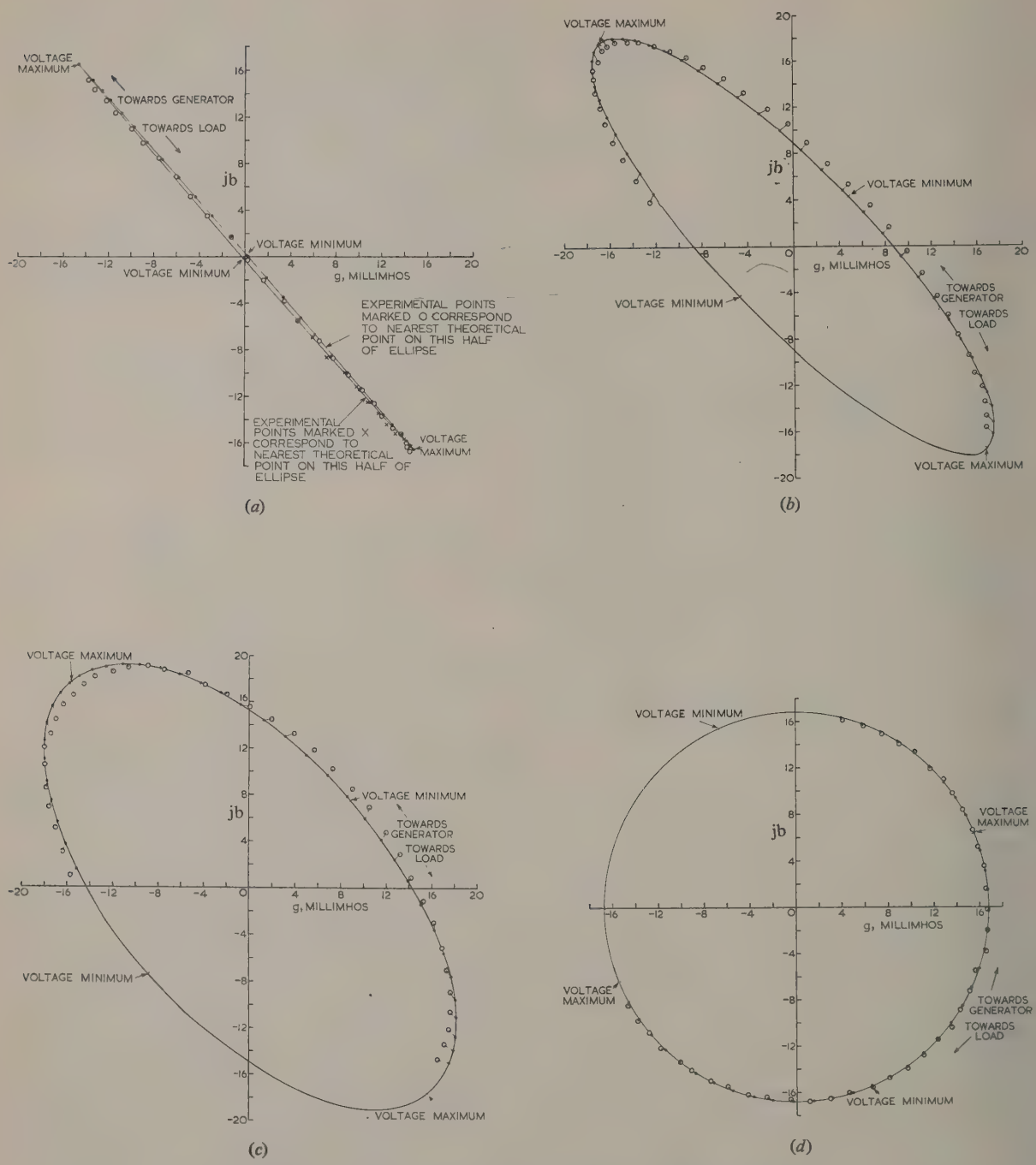
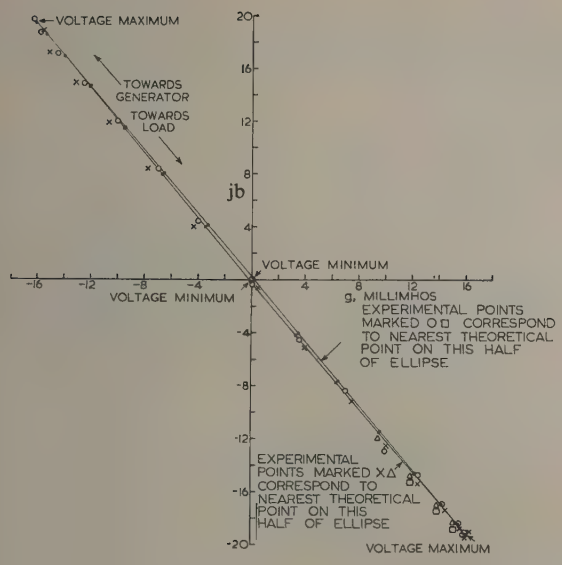


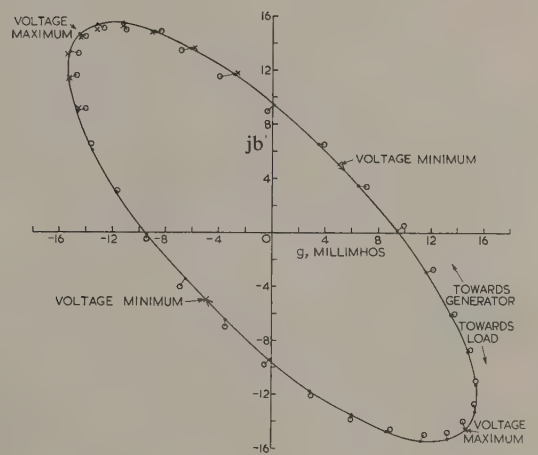
Fig. 4.—Complex voltage distribution along a slotted line at 500 Mc/s.

(*a*) S.W.R. = 0·007.
(*b*) S.W.R. = 0·263.
(*c*) S.W.R. = 0·5.
(*d*) S.W.R. = 0·98.

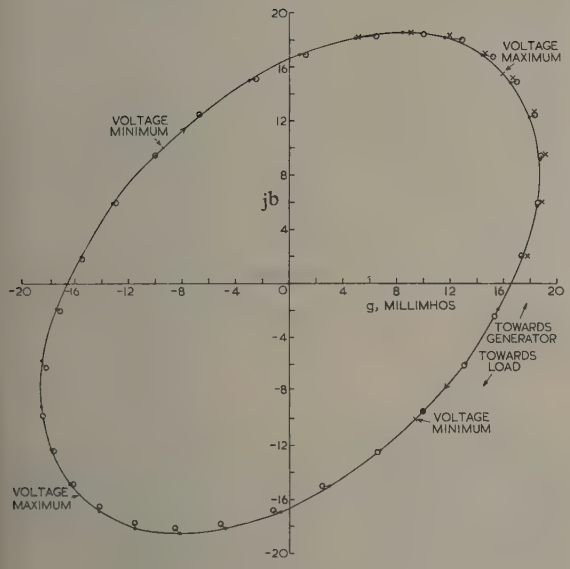
•——• Theoretical.
○ × Experimental.



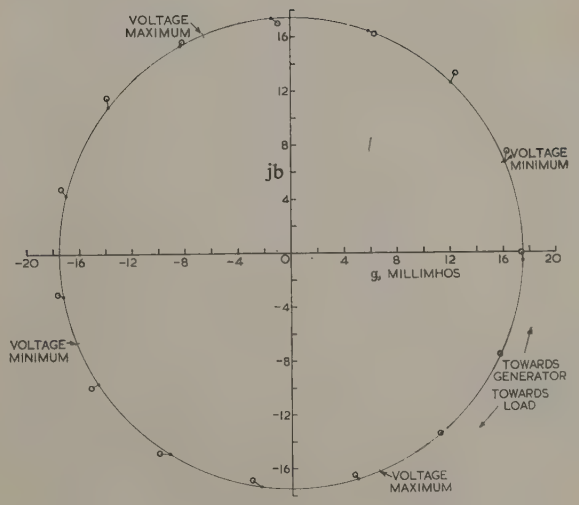
(*a*)



(*b*)



(*c*)



(*d*)

Fig. 5.—Complex voltage distribution along a slotted line at 1 000 Mc/s.

(*a*) S.W.R. = 0·007.
(*b*) S.W.R. = 0·34.
(*c*) S.W.R. = 0·62.
(*d*) S.W.R. = 0·98.

•——• Theoretical.
□○×△ Experimental: different symbols are used to distinguish points in different cycles [in (*a*), different half-cycles].

over a wide range. The meter scales are marked in approximately equal intervals of amplitude, which makes accurate reading difficult at the minima. Phase measurement errors are therefore greatest where the magnitude is a minimum. The accuracy of phase measurement at 500 and 1 000 Mc/s is better than $\pm 3°$ if the range of attenuation is within 6 dB, but, owing to lack of scale discrimination, errors up to $\pm 10°$ may occur at the minima if the range is extended to 40 dB.

A convenient method of expressing the accuracy is in terms of the vector error. If δg and δb are the errors in the readings of the real and imaginary parts, obtained from the conductance and susceptance scales of the admittance meter, the magnitude of the vector error is $(\delta g^2 + \delta b^2)^{1/2}$. It corresponds to the length of a line drawn on the Argand diagram between theoretical and experimental points. If the maximum readings of the scales are g_m and b_m, the relative vector error, ϵ, may be defined as

$$\epsilon = (\delta g^2 + \delta b^2)^{1/2}/(g_m^2 + b_m^2)^{1/2}$$

The maximum value of ϵ is approximately 6% for the frequencies and standing-wave ratios used in the experiments.

The greater part of the error is associated with the difference (5–10%) in eccentricity between the measured and theoretical ellipses. This discrepancy may be due to either admittance meter errors or imperfect screening of the apparatus.

A large difference (many wavelengths) between the lengths of the two paths from oscillator to receiver should be avoided, otherwise the accuracy of measurement will be very dependent on the frequency stability of the signal source.

(5) CONCLUSIONS

The accuracy of phase measurement is better than $\pm 3°$ unless large variations of amplitude are encountered, in which case errors of scale reading at the minima limit the accuracy of phase measurement. If the error is regarded as a vector which joins the true and measured points when plotted on the Argand diagram, the magnitude of this error vector does not exceed about 6% of the full-scale reading of the instrument.

The method of measurement described has been used for measuring the radiation patterns of model aerial systems, the 'network' consisting of the aerial under test, which could be rotated, a fixed aerial and the space between them. A knowledge of the phase variation is often valuable, making it possible to work with a very simple model. For example, where a v.h.f. transmitting aerial consisting of rings of radiating dipoles surrounding a supporting mast is required to have a horizontal radiation pattern of a particular shape, a model mast section is arranged to carry only one dipole. The radiation pattern is then measured in amplitude and phase, and from this is calculated the radiation pattern of a complete ring, with the dipoles energized in any manner. In this way the best choice of the amplitudes and phases of the feeds to the dipoles may be determined by calculation. One advantage of this technique is that the dipole used need not be well matched to the feeder. If a complete ring of elements were used for the measurement, great care in matching would be required to achieve the required amplitudes and phases of the radiating currents.

Simultaneous phase and amplitude measurement is also useful when adjusting directional couplers, where the effect of imperfect directivity has to be distinguished from that of an imperfectly matched load.[2]

(6) ACKNOWLEDGMENTS

The authors are indebted to the Chief Engineer of the British Broadcasting Corporation for permission to publish the paper. Thanks are due to the General Radio Company, Cambridge, Massachusetts, U.S.A., for permission to reproduce Fig. 2.

(7) REFERENCES

(1) ROBERTSON, S. D.: 'A Method of Measuring Phase at Microwave Frequencies', *Bell System Technical Journal*, 1949, **28**, p. 99.
(2) SOSIN, B. M.: 'Application of Directional Couplers to Amplitude Ratio and Phase Measurement at Very High Frequencies', *Marconi Review*, 1st Quarter, 1951, p. 39.
(3) THURSTON, W. R.: 'A Transadmittance Meter for V.H.F.-U.H.F. Measurements', *Institute of Radio Engineers Convention Record*, 1956, Part 5, p. 3.
(4) *General Radio Experimenter*, May, 1950, and August, 1953.

(8) APPENDIX: THE DISTRIBUTION OF VOLTAGE ON A UNIFORM TRANSMISSION LINE

The voltage on a uniform transmission line may be regarded as the sum of two components, associated respectively with waves travelling in two directions. Neglecting loss, and assuming only one frequency to be present, the voltage V may be expressed as

$$V = V_i \varepsilon^{j\beta l} + \rho V_i \varepsilon^{-j\beta l} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where βl is the angular distance measured along the line from an arbitrarily chosen origin towards the source, V_i is the voltage associated with the incident wave at the origin, and ρ is the reflection coefficient at the origin. In general ρ will be complex.

Let $$\rho = |\rho|\varepsilon^{j\phi}$$

Then

$$\begin{aligned} V &= V_i \varepsilon^{j\phi/2}[\varepsilon^{j(\beta l - \phi/2)} + |\rho|\varepsilon^{-j(\beta l - \phi/2)}] \\ &= V_i \varepsilon^{j\phi/2}[(1 + |\rho|)\cos(\beta l - \phi/2) + j(1 - |\rho|)\sin(\beta l - \phi/2)] \end{aligned}$$

. . . . (2)

Now the parametric equations of an ellipse centred at the origin, with the major and minor axes (lengths $2a$ and $2b$) oriented along the x and y axes are

$$\left.\begin{aligned} x &= a\cos\theta \\ y &= b\sin\theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Writing $z = x + jy$, the equation of the ellipse as a locus on the Argand diagram is

$$z = a\cos\theta + jb\sin\theta \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Comparing eqn. (4) with eqn. (2) it is seen that, if the quantity in square brackets is plotted on the Argand diagram, the locus traced is an ellipse with major and minor semi-axes $(1 + |\rho|)$ and $(1 - |\rho|)$ on the real and imaginary axes of the z-plane. The ratio of the axes, $(1 - |\rho|)/(1 + |\rho|)$, is equal to the voltage standing-wave ratio. When V is plotted, this ellipse is scaled in proportion to V_i and rotated through the angle $\phi/2$.

621.376.3

The Institution of Electrical Engineers
Paper No. 3182 E
Mar. 1960
©

FUNDAMENTAL AND HARMONIC DISTORTION OF WAVES FREQUENCY-MODULATED WITH A SINGLE TONE

By R. G. MEDHURST, B.Sc.

(*The paper was first received* 30*th April, and in revised form* 16*th October*, 1959.)

SUMMARY

The distortion of single-tone frequency modulation by passive networks whose amplitude and phase characteristics are non-linear with frequency is considered in some detail. A new approximate formula has been developed by extension of earlier methods, and its use, supplemented by the quasi-stationary approximation and by direct Fourier analysis, permits evaluation of distortion levels over ranges of parameters previously excluded. A number of examples are given, comparison being made where possible with results obtained by other methods. Particular attention is given to distortion of the amplitude and phase of the fundamental, this being of importance in connection with television baseband requirements. An attempt is made to clarify certain theoretical aspects, such as the validity of the 'quasi-stationary' solution, which have been the subject of controversy.

LIST OF PRINCIPAL SYMBOLS

ω_c = Carrier frequency, rad/sec.
ω_M = Frequency modulation, rad/sec.
ω_a = Modulation frequency, rad/sec.
ω_d = Frequency deviation, rad/sec.
β = Modulation index, $=\omega_d/\omega_a$.
μ_t = Phase modulation, rad.
$1 + \rho(\omega)$ = Amplitude transmission characteristic at frequency ω (voltage ratio).
$\phi(\omega)$ = Phase transmission characteristic at frequency ω, rad.
Δ = Phase-modulation distortion, rad.
Δ_1 and Δ_2 = Distortion terms defined by expression (3).
$1 + \Delta_3$ = Amplitude modulation imposed by the distorting network (voltage ratio).
V = Amplitude of the modulated wave before passage through the distorting network, volts.
J_n = A Bessel function of the first kind of order n and argument β.
ω_B = Semi-bandwidth of a maximally-flat-amplitude triple-tuned circuit, between half-power points, rad.
$p = \omega_a/\omega_B$.
$q = \omega_d/\omega_B$.

(1) INTRODUCTION

When a frequency-modulated wave traverses a passive network whose amplitude and phase response will, in general, vary non-linearly with frequency, the output wave is found to contain both frequency-modulation distortion and amplitude modulation. In principle, both of these forms of distortion could be evaluated by first modifying each sideband of the incoming wave in accordance with the transfer characteristic of the network at the particular sideband frequency, and then rewriting the resulting spectrum as a simultaneously amplitude- and frequency-modulated carrier. The first step can be carried out without excessive difficulty for single-tone modulation, provided that the modulation index is not very large. However, if the modulation contains more than one tone, the associated spectrum becomes analytically so complex as to be outside the range of economic computation, except for a few special cases.[1–4] The second step usually involves formidable difficulties, and recourse must be had to approximate methods, often with an unknown associated error.

Faced with this situation, a number of authors (following Carson and Fry[5]) have attempted to operate with the transmission characteristic directly on the modulated carrier, thus avoiding the intermediate step of expansion into a spectrum. An immediate restriction on this approach is that it becomes necessary to express the network characteristics in a, preferably simple, analytical form: even within this limitation, substantial analytical difficulties arise.

Not surprisingly, this problem has given rise to a considerable body of literature during the last 20 years or so. Some of the more recent work is given in References 6–16; Brown[13] presents a valuable survey of earlier results, with particular attention to single-tone modulation.

Single-tone modulation has been extensively investigated, perhaps to some extent because of all cases this makes the nearest approach to analytical simplicity. However, it does not seem possible to draw useful conclusions regarding the behaviour of systems containing more complicated modulating signals, such as multiplex telephony, from results relating to single-tone modulation. Thus, the tendency in recent times has been to work with more realistic signals, such as band-limited white noise.

With increasing interest in the transmission of both black-and-white and colour television through f.m. trunk radio systems, it becomes of value to reconsider the distortion of single-tone modulation, with particular reference to fundamental distortion. This is because one of the specifications relating to television transmission concerns the amplitude and phase variation with frequency of a single input modulating tone[17] (i.e. the 'steady-state' response).

In the present paper an earlier first-order theory[8, 12] is extended to make it of more general use. Applications are made to various numerical cases, including some previously treated by different methods. It is shown that use of the new formula, supplemented by the quasi-stationary result and numerical Fourier analysis, can give distortion levels over a fuller range of parameters than has been hitherto possible. In addition, consideration is given to some aspects of frequency-modulation theory which have caused difficulty in earlier work.

(2) APPROXIMATE FIRST-ORDER THEORY

What is referred to in the present paper as the 'approximate first-order theory' was developed in References 8 and 13. This technique has the valuable feature that the transmission characteristics need be known only at the discrete sideband frequencies, so that, for example, discontinuous or experimental characteristics can be dealt with. The various expansion

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Ltd., Wembley, England.

formulae,[5, 7, 10] on the other hand, require that the characteristics be known in analytical form, and are easily applicable only to fairly low-order characteristics. The derivation of the approximate first-order theory will be recalled briefly in the present Section.

The undistorted wave is

$$V \sin(\omega_c t + \mu_t) = V \cos \mu_t \sin \omega_c t + V \sin \mu_t \cos \omega_c t \quad . \quad (1)$$

After passage through the distorting network, this becomes

$$\begin{aligned} V(1 + \Delta_3) \sin(\omega_c t + \mu_t + \Delta) \\ = V(1 + \Delta_3)(\cos \mu_t - \Delta \sin \mu_t) \sin \omega_c t \\ + V(1 + \Delta_3)(\sin \mu_t + \Delta \cos \mu_t) \cos \omega_c t \quad . \quad (2) \end{aligned}$$

to first order in Δ.

The right-hand side of eqn. (1) can, in principle, be developed into the r.f. spectrum of the undistorted wave. Suppose that after passage through the distorting network the spectrum takes the form

$$V(\cos \mu_t + \Delta_1) \sin \omega_c t + V(\sin \mu_t + \Delta_2) \cos \omega_c t \quad . \quad (3)$$

where Δ_1 and $\Delta_2 \ll 1$. Then, from expressions (2) and (3),

$$\begin{aligned} (1 + \Delta_3)(\cos \mu_t - \Delta \sin \mu_t) &= \cos \mu_t + \Delta_1 \\ (1 + \Delta_3)(\sin \mu_t + \Delta \cos \mu_t) &= \sin \mu_t + \Delta_2 \end{aligned}$$

Eliminating Δ_3 and retaining only first-order terms in the Δ's, it is found that

$$\Delta = \Delta_2 \cos \mu_t - \Delta_1 \sin \mu_t \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Using this result and considering single-tone modulation of the form $\mu_t = \beta \sin \omega_a t$, it is shown[8] that the Nth harmonic component of the phase-modulation distortion is given by

$$\begin{aligned} \cos(N\omega_a t) \sum_{n=-\infty}^{\infty} J_n(J_{n-N} + J_{n+N}) \phi(\omega_c + n\omega_a) \\ + \sin(N\omega_a t) \sum_{n=-\infty}^{\infty} J_n(J_{n-N} - J_{n+N}) \rho(\omega_c + n\omega_a) \quad . \quad (5) \end{aligned}$$

In developing this expression it was assumed that both ϕ and ρ are very much less than unity over the range of significant sidebands. With $N = 1$ in expression (5) it will be noticed that, to this approximation, only the phase characteristic produces phase distortion of the fundamental. The coefficients which multiply the characteristic values have been tabulated[8] and plotted in more detail[13] for $N = 2$ and 3.

(3) EXTENDED FIRST-ORDER THEORY

The condition that both ϕ and ρ should be small, which is required for expression (5) to be valid, is, in fact, a rather serious restriction on the application of the formula. Common practice is to design networks so that ϕ and ρ are small over the frequency range within which sideband amplitudes are substantial, allowing rapid deterioration outside this range. Numerical results obtained from expression (5) would suggest—incorrectly—that characteristics must be tightly controlled over a considerably wider band than is usually contemplated.

The range of application of the first-order theory can be considerably extended when it is realized that two quite different sorts of approximation are incorporated in expression (5), only one of which is actually necessary in a first-order approach. One approximation, which is the basic first-order assumption, enters the theory quite early: it occurs when second- and higher-order terms in the Δ's are neglected in the derivation of eqns. (2) and (4). The other approximation consists in writing $\cos \phi = 1$ and $\sin \phi = \phi$, on the assumption that ϕ is small over the full range of significant sidebands. The $\cos \phi$ and $\sin \phi$ can, in fact, be retained without leading to excessive complication. The analysis (Section 10.1) follows a similar course to that in Reference 8. It is found that the Nth harmonic component of the phase modulation distortion is now given by

$$\begin{aligned} \cos(N\omega_a t) \sum_{n=-\infty}^{\infty} J_n(J_{n-N} + J_{n+N}) \\ \{[1 + \rho(\omega_c + n\omega_a)] \sin[\phi(\omega_c + n\omega_a)]\} \\ + \sin(N\omega_a t) \sum_{n=-\infty}^{\infty} J_n(J_{n-N} - J_{n+N}) \\ \{[1 + \rho(\omega_c + n\omega_a)] \cos[\phi(\omega_c + n\omega_a)] - 1\} \\ \quad . \quad . \quad . \quad . \quad (6) \end{aligned}$$

When ϕ is sufficiently small and the product of ρ and ϕ is neglected, expression (6) reduces, correctly, to expression (5).

The extended theory embodied in expression (6) has a wider range of application than the approximate theory, while still remaining valid as a first-order theory, because in certain circumstances the Δ's remain small when ϕ is not small.

(4) THE MAXIMALLY-FLAT-AMPLITUDE TRIPLE-TUNED CIRCUIT (EXAMPLE 1)

For the first example of the use of expression (6) we consider a transmission network of considerably practical use, namely the maximally-flat-amplitude triple-tuned circuit.[18] This was treated in some detail in References 9 and 10. Two methods of approach were used, one valid for high values of β and one for low. For intermediate regions, the amplitude and phase distortion of the fundamental had to be estimated by graphical interpolation, and, in fact, for too large a normalized frequency deviation (denoted here by q and in References 9 and 10 by Δ) it becomes impossible to bridge the gap plausibly (see Figs. 5 and 6 of Reference 10). In the present Section it will be shown that complete coverage can be obtained by judicious use of the extended first-order theory, quasi-stationary theory and direct Fourier analysis.

Taking the mid-band frequency as being equal to the carrier frequency of the incoming wave, the amplitude and phase characteristics are:[18]

$$1 + \rho(\omega) = \frac{1}{\sqrt{(1 + X^6)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\phi(\omega) = -\arctan\left(\frac{2X - X^3}{1 - 2X^2}\right) \quad . \quad . \quad . \quad (8)$$

where

$$X = Q\left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right)$$

We shall consider the high-Q case, so that $Q \simeq \omega_c/2\omega_B$. Following Collins and Skwirzynski,[9, 10] both the modulating frequency and the frequency deviation will be normalized with respect to ω_B. Then, for insertion into expression (6), characteristic values are required at frequencies $\omega_c + p\omega_B$. On the high-Q assumption, it is easy to show that $X \simeq np$, at the sideband frequency $\omega_c + np\omega_B$.

To illustrate the method, a normalized deviation, q, equal to unity will be taken. This is the largest value of q considered in Reference 10, and will be used here because interpolation of the curves in Reference 10 is most difficult for this value, and consequently the advantage of the present approach shows up most clearly.

For a fixed value of q the extended first-order theory is applicable only for an intermediate range of the normalized modulating frequency, p. When p is very small, the number

of terms of expression (6) required becomes prohibitively large. It is then necessary to resort to the quasi-stationary theory (see Section 5). When p becomes large, the distortion introduced by the circuit becomes so great as to invalidate a first-order approach. In this region, direct Fourier analysis can be used. Since the treatment of this example is of general application, it is worth considering the three methods separately and in some detail.

(4.1) Quasi-Stationary Theory

Before applying either the quasi-stationary or the extended first-order theories it will be advisable to remove the linear part of $\phi(\omega)$ (defined in terms of the first derivative at carrier frequency), so that the phase characteristic to be inserted into expression (6) becomes

$$\phi(\omega) = -\arctan\left(\frac{2X - X^3}{1 - 2X^2}\right) + 2X \quad . \quad . \quad (9)$$

the wave being considered subsequently to pass through a correcting network having a phase characteristic

$$\phi(\omega) = -2X \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The validity of this operation is discussed in Section 7. The effect of eqn. (10) is to produce a delay on the modulation terms in consequence of which an Nth harmonic in the modulation will undergo a phase shift $-2Np$ (Section 7). Thus distortion components evaluated using expression (9) must be phase corrected by this amount.

Quasi-stationary theory neglects the amplitude characteristic of the distorting network and assumes that the carrier phase follows the phase characteristic as the carrier frequency varies. Writing the undistorted wave in the form

$$\sin\left[\omega_c t + \frac{q}{p}\sin(p\omega_B t)\right]$$

the frequency modulation is

$$q\omega_B \cos(p\omega_B t)$$

so that, from eqn. (9), the distorted phase modulation becomes

$$\frac{q}{p}\sin(p\omega_B t) - \arctan\left[\frac{2q\cos(p\omega_B t) - q^3\cos^3(p\omega_B t)}{1 - 2q^2\cos^2(p\omega_B t)}\right] + 2q\cos(p\omega_B t)$$

Putting $q = 1$ and writing the powers of cosines as multiple angles, this is

$$\frac{1}{p}\sin(p\omega_B t) + \arctan\left[\frac{5\cos(p\omega_B t) - \cos(3p\omega_B t)}{4\cos(2p\omega_B t)}\right] + 2\cos(p\omega_B t)$$

The coefficient of the fundamental portion of the inverse tangent is given by

$$\frac{2}{\pi}\int_0^{\pi} \arctan\left[\frac{5\cos(p\omega_B t) - \cos(3p\omega_B t)}{4\cos(2p\omega_B t)}\right]\cos(p\omega_B t)\,d(p\omega_B t) = -2{\cdot}2919$$

by numerical integration. Thus the fundamental in the phase modulation becomes

$$\frac{1}{p}\sin(p\omega_B t) - 0{\cdot}2919\cos(p\omega_B t)$$

and the fundamental phase shift is

$$-\arctan(0{\cdot}2919p) - 2p \quad . \quad . \quad . \quad (11)$$

the second term being the correction required to compensate for the removal of the linear term from the original phase characteristic.

(4.2) Direct Fourier Analysis

If the number of significant sidebands is not too large (i.e. in the low-modulation-index case), the f.m. distortion components can be evaluated directly. Writing the modulated wave (with, for convenience, unit amplitude) in the form

$$\sin(\omega_c t + \beta\sin\omega_a t) \quad . \quad . \quad . \quad . \quad (12)$$

the corresponding spectrum is

$$\sum_{n=-\infty}^{\infty} J_n(\beta)\sin[(\omega_c + n\omega_a)t]$$

which, after distortion, becomes

$$\sum_{n=-\infty}^{\infty} [1 + \rho(\omega_c + n\omega_a)]J_n(\beta)\sin[(\omega_c + n\omega_a)t + \phi(\omega_c + n\omega_a)] \quad (13)$$

To extract the phase-modulation term from expression (13), all that is necessary is to rewrite the expression in the form $A\cos\omega_c t + B\sin\omega_c t$, so that the phase modulation appears in the form arc tan A/B. Specifically, it is found that the distorted phase modulation is

$$\arctan\left(\frac{2\sum_{n=1}^{\infty}\{1 + \rho[\omega_c + (2n-1)\omega_a]\}J_{2n-1}(\beta)\sin\{(2n-1)\omega_a t + \phi[\omega_c + (2n-1)\omega_a]\}}{[1 + \rho(\omega_c)]J_0(\beta) + 2\sum_{n=1}^{\infty}[1 + \rho(\omega_c + 2n\omega_a)]J_{2n}(\beta)\cos[2n\omega_a t + \phi(\omega_c + 2n\omega_a)]}\right) \quad . \quad . \quad . \quad . \quad (14)$$

when $\rho(\omega)$ and $\phi(\omega)$ are respectively even and odd functions about $\omega = \omega_c$. In Reference 9 an expression equivalent to (14) is evaluated by a contour-integration method. However, the number of sidebands which can be dealt with in this way is quite limited, and if access can be had to a digital computer, it is preferable to extract fundamental and harmonics from expression (14) by a numerical Fourier analysis.[19, 20]

For the particular circuit and circuit parameters considered in the present Section, modulation indices up to two can be comfortably handled, the highest-order sideband considered when $\beta = 2$ being the sixth. The most straightforward method of deciding whether sufficient sidebands have been taken in expression (14) is to add a further one, re-evaluate expression (14) and compare the revised fundamental and harmonics with the original. A more economical procedure is to notice that neglect of the rth sideband amounts to assuming that $\rho(\omega_c + r\omega_a) = -1$, instead of its actual value. From the theory of Section 3, the approximate effect on the Nth harmonic of the phase modulation due to neglecting the rth sideband can be found by subtracting from the phase modulation the expression

$$J_r(J_{r-N} - J_{r+N})\{[1 + \rho(\omega_c + r\omega_a)]\cos[\phi(\omega_c + r\omega_a)] + (-1)^{N+1}[1 + \rho(\omega_c - r\omega_a)]\cos[\phi(\omega_c - r\omega_a)]\}\sin(N\omega_a t)$$

$$+ J_r(J_{r-N} + J_{r+N})\{[1 + \rho(\omega_c + r\omega_a)] \sin [\phi(\omega_c + r\omega_a)] + (-1)^N[1 + \rho(\omega_c - r\omega_a)] \sin [\phi(\omega_c - r\omega_a)]\} \cos (N\omega_a t) \quad . \quad . \quad . \quad . \quad (15)$$

In carrying out a direct Fourier analysis there is, of course, no necessity to remove initially the linear portion of the phase characteristic.

(4.3) Extended First-Order Theory

As in the quasi-stationary theory, we first remove the linear portion of the phase characteristic, making a subsequent correction to the phases of all the phase-modulation components. Computation from expression (6) is now straightforward, although an increasing number of terms must be considered as β increases. For the circuit and circuit parameters considered in the present Section, β values up to 9 were fairly easily manageable using a desk calculator.

(4.4) Numerical Results

Amplitude and phase responses of the fundamental are shown in Figs. 1 and 2 as functions of normalized modulation frequency when the normalized frequency deviation is unity. The phase

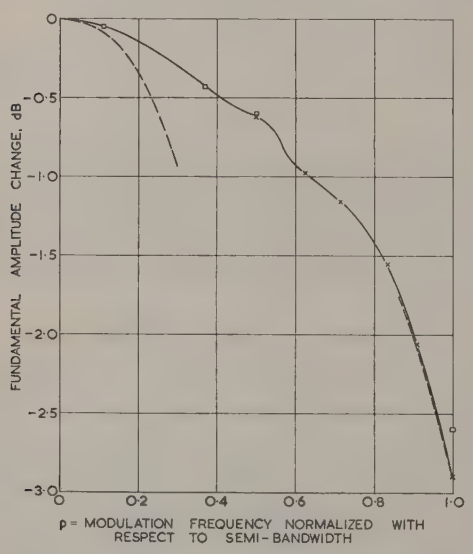


Fig. 1.—Amplitude response of the fundamental for the maximally-flat-amplitude triple-tuned circuit (normalized frequency deviation, q, $=1$).

× Numerical Fourier analysis.
⊡ Extended first-order theory.
—— From Fig. 5 of Reference 10.

shift is plotted as departure from the 'static' phase shift, as in Reference 10. It has been shown by Skwirzynski[9] that under certain circumstances an approximation to the fundamental phase shift, when the transmission amplitude characteristic is symmetrical and the transmission phase characteristic is skew-symmetrical about the carrier frequency, is given by the transmission phase-characteristic value at the first upper sideband frequency, this being called the 'static' phase shift (Borodich[7] gives an extension of this result for general characteristics). This appears to be an approximation valid only for low-order characteristics: it will obviously not be true, for example, for a phase characteristic which is flat to the first upper and lower sideband frequencies and varies rapidly there-

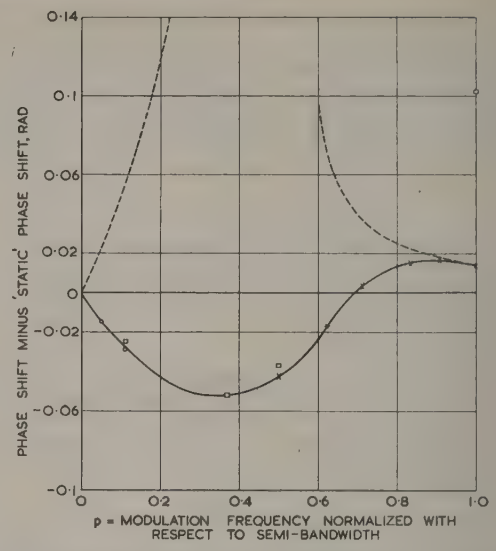


Fig. 2.—Relative phase response of the fundamental for the maximally-flat-amplitude triple-tuned circuit (normalized frequency deviation, q, $=1$).

× Numerical Fourier analysis.
⊡ Extended first-order theory.
○ Quasi-stationary theory.
—— From Fig. 6 of Reference 10.

after. However, it is seen from Fig. 2 to hold rather well for the present case.

The solid curves in both Figs. 1 and 2 have been drawn to pass through the exact points obtained by Fourier analysis; the extended first-order theory and (in Fig. 2) the quasi-stationary approximation were used to carry the curves down to $p = 0$. In Fig. 2, as $p \to 1$ the result from the extended first-order theory departs considerably from the true value. This is because (as appears from Fig. 1) in this region it is no longer accurate to assume that the distortion level is small, so that first-order theory here breaks down.

For comparison, curves from Figs. 5 and 6 of Reference 10 are reproduced in Figs. 1 and 2, terminating at the estimated limit of reliability of the respective formulae. It is apparent that for the intermediate ranges of p no satisfactory graphical interpolation is possible. In Fig. 2 it is seen that there is a disagreement of sign between the curve from Reference (10) and the present curve. It appears, in fact, that there are two errors of sign in the analysis of Reference 10, one of which applies to both the high- and low-p expressions, and the other, cancelling the first, only to the high-p expression.

(5) THE QUASI-STATIONARY APPROXIMATION

(5.1) General Discussion

There has been considerable divergence of opinion in the literature as to the conditions of validity of the quasi-stationary approximation (see, for example, References 10 and 13). At one extreme the suggestion has been made that quasi-stationary conditions may not be approached, even in the limit of very large modulation index. Indeed, for a particular network, curves of second- and third-harmonic distortion have been shown[13] in which the quasi-stationary solution departs increasingly from levels obtained by the approximate first-order theory (Section 2) as β increases.

The statement of the quasi-stationary approximation is, in fact,

merely a rephrasing of the definition of a phase/frequency transmission characteristic. The phase/frequency characteristic is obtained by passing through the network single-frequency tones of various frequencies, in each case measuring the phase shift after sufficient time has elapsed to allow transients to die out. In the quasi-stationary approximation this is made a continuous process, with the proviso that the variation of the frequency shall be so gradual that transients have no measurable effect on the phase variations. In particular, if the frequency varies slowly in a sinusoidal manner, the additional carrier phase variation is then to be obtained by inserting this sinusoid into the phase transmission characteristic. All that remains to be decided is how slowly the input frequency must vary in order that the phase of the carrier shall follow the phase transmission characteristic. This is, of course, the central problem; it will depend on individual circumstances, and an answer may not be easy to give. However, the quasi-stationary approximation must always be a valid ultimate limit; if it appears not to be so, one must look for an inadequacy in the analysis.

(5.2) Minimum-Phase Network having Linear Amplitude Slope (Example 2)

Consideration of the behaviour of this network is particularly instructive, since, according to an earlier analysis, it appeared to violate the requirement that quasi-stationary conditions should be approached in the limit of large values of β.

The amplitude characteristic is supposed to be constant for all frequencies from zero up to $\omega_c + \omega_\Delta$, and to increase thereafter at the constant rate of k dB/rad/sec, levelling off again at a frequency beyond the range of significant sidebands. Using the minimum phase condition, an approximate form for the phase characteristic, derived in Reference 13, is

$$\phi(\omega_c + \omega) = \frac{k\omega}{8{\cdot}686\pi}\left[1 + \left(\frac{\omega_\Delta}{\omega} - 1\right)\log\left|1 - \frac{\omega}{\omega_\Delta}\right|\right] \quad (16)$$

In deriving this expression, the linear term (see Section 7) was subtracted initially. According to the quasi-stationary approximation, the outcoming phase modulation is obtained by substituting for ω the frequency modulation, which, for sinusoidal modulation, is $\beta\omega_a \cos(\omega_a t)$. Thus the phase modulation becomes

$$\beta \sin \omega_a t + \frac{k}{8{\cdot}686\pi}\left\{\beta\omega_a \cos(\omega_a t) + [\omega_\Delta - \beta\omega_a \cos(\omega_a t)] \log\left|1 - \frac{\beta\omega_a}{\omega_\Delta}\cos\omega_a t\right|\right\} \quad (17)$$

Fundamental and harmonics have to be extracted from expression (17) by evaluating the appropriate Fourier integrals, the course of the analysis depending on whether $\beta\omega_a/\omega_\Delta$ is less or greater than unity. In the latter case, some care has to be taken in integrating through a singularity. It is found (Section 10.2) that the ratio of second-harmonic amplitude to fundamental amplitude in the frequency modulation is given by one or other of the following expressions:

$$\frac{\text{Second harmonic level}}{\text{Fundamental level}} = \frac{k\omega_a}{4{\cdot}343\pi}$$

$$\left[\frac{3K^2 - 2}{3K^3} + \frac{2(1 - K^2)^{3/2}}{3K^3}\right] \text{(voltage ratio, } K < 1)$$

$$= \frac{k\omega_a}{4{\cdot}343\pi}\left(\frac{3K^2 - 2}{3K^3}\right) \text{(voltage ratio, } K > 1) \quad (18)$$

where $K = \beta\omega_a/\omega_\Delta$

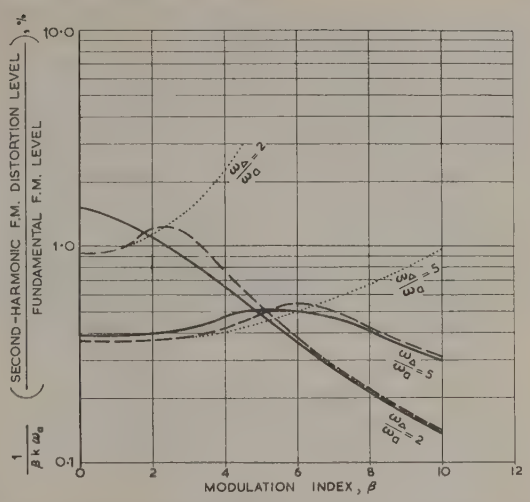


Fig. 3.—Second-harmonic distortion due to minimum-phase response associated with a linear-amplitude slope: comparison of approximate first-order and quasi-stationary theories.

—— Approximate first-order theory.[8]
– – Quasi-stationary theory.
···· Quasi-stationary theory according to Reference 13.

In Fig. 3, second-harmonic distortion levels evaluated from eqn. (18) are plotted for two representative values of ω_Δ/ω_a, together with second-harmonic levels evaluated according to the approximate first-order theory outlined in Section 2 (which may be used instead of the extended first-order theory so long as k is small). It is seen, as expected, that with increasing modulation index the two curves approach each other. In this Figure are also plotted the quasi-stationary values given in Reference 13. They are based on an expansion of the logarithmic term in expression (17) which is actually valid only for $K < 1$. The reason for the departure of these curves from the approximate first-order curves with increasing values of β is that this expansion has been used outside its range of validity.

(5.3) Limits of the First-Order Theories with Increasing β

Consideration of the validity of the quasi-stationary solution led to an investigation of the limiting behaviour of the first-order theories as β increases. The remarkable result has emerged that the approximate first-order expression (5) takes a limiting form, for large values of β, which is, in fact, the expression which follows from the quasi-stationary approach (Section 10.3). This is the more surprising in that the approximate first-order theory is based on the assumption that the amplitude- and phase-characteristic variations are small [and, as will be seen in Figs. 4(*a*) and (*b*), breaks down badly when this condition is violated], whereas the quasi-stationary solution holds as a limiting form for any size of characteristic variation. The limit of the extended first-order theory can be obtained in a form analogous to the quasi-stationary solution. It is, however, sufficiently more complicated to make evaluation of the harmonics prohibitively difficult for most shapes of characteristics. In practice, when β is sufficiently large for limiting conditions to be approached, the quasi-stationary form will normally be accurate enough.

(6) FIFTH-ORDER PHASE CHARACTERISTIC (EXAMPLE 3)

The superiority of the extended first-order theory over the approximate first-order theory shows up clearly when one considers phase characteristics varying slowly over the range of

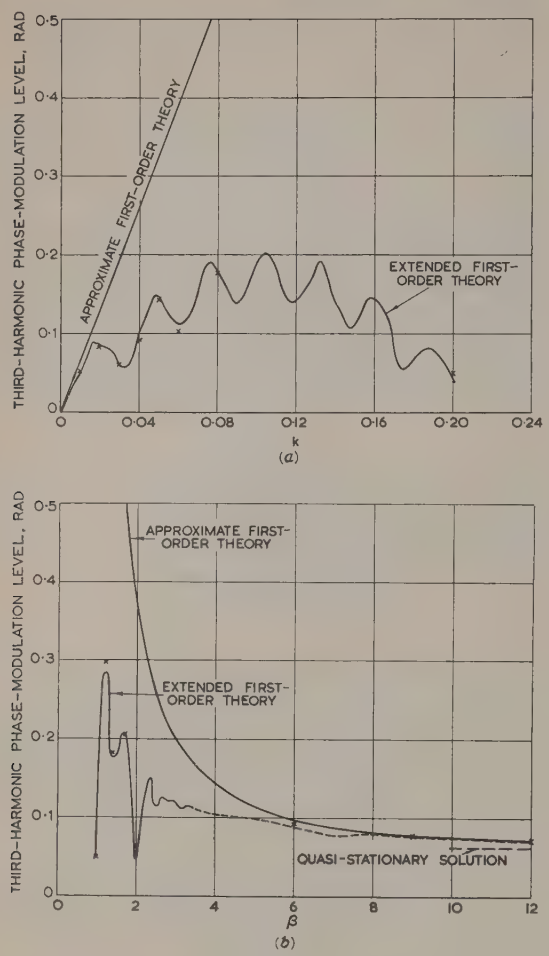


Fig. 4.—Third-harmonic distortion due to a phase characteristic of the form $\phi(\omega - \omega_c) = K\left(\frac{\omega - \omega_c}{\omega_d}\right)^5$.

× Numerical Fourier analysis.
(*a*) K variable, $\beta = 1$.
(*b*) $K = 0{\cdot}2$, β variable.

important sidebands and rapidly outside. As an example, Figs. 4(*a*) and (*b*) show the behaviour of third-harmonic distortion due to a fifth-order phase characteristic, the amplitude characteristic being flat. The form assumed for the phase characteristic is

$$\phi(\omega - \omega_c) = K\left(\frac{\omega - \omega_c}{\omega_d}\right)^5 \quad . \quad . \quad . \quad . \quad (19)$$

Taking the phase modulation as $\beta \sin \omega_a t$, so that the frequency modulation is $\omega_d \cos \omega_a t$, the quasi-stationary solution for the phase-modulation distortion is

$$K \cos^5 \omega_a t = \frac{5}{8} K \cos \omega_a t + \frac{5}{16} K \cos 3\omega_a t + \frac{1}{16} K \cos 5\omega_a t \quad (20)$$

Fig. 4(*a*) shows how the third-harmonic phase modulation varies with K, for a fixed value of β, according to the approximate and extended first-order theories. As K increases, the error in the approximate first-order theory becomes rapidly larger. Furthermore, the fine structure which appears in the curve corresponding to the more exact theory is lost in the approximate theory. Some points obtained by direct Fourier analysis are in good agreement with the more exact theory.

Fig. 4(*b*) shows corresponding curves for $K = 0{\cdot}2$ and varying β. The curve for the extended first-order theory is shown dotted for β values above 3·6, because insufficient points were computed in this region to determine the detailed fine structure. Again there is good agreement between the more exact theory and values obtained by direct Fourier analysis. Both the first-order curves approach the quasi-stationary result, given by the second term of expression (20), as β increases.

(7) CONSTANT AND LINEAR TERMS IN THE PHASE CHARACTERISTIC

Although the theory of frequency-modulation distortion by general characteristics involves substantial difficulties, the special cases of constant and linear phase characteristics can be solved exactly (as can those of constant and linear amplitude characteristics, e.g. Reference 12). Consider a typical sideband of the form $a_n \sin [\omega_c t + n\omega_a t + \phi_n]$ and apply to it a phase characteristic of the form $\psi + \tau(\omega - \omega_c)$. The sideband becomes $a_n \sin [\omega_c t + n\omega_a t + \phi_n + \psi + \tau n\omega_a]$, i.e. $a_n \sin [(\omega_c t + \psi) + n\omega_a(t + \tau) + \phi_n]$. It is apparent that, when the sidebands are recombined to form a modulated carrier, the carrier phase will be changed by an amount ψ, and all components of the modulation delayed by an amount $-\tau$, so that the effect of the characteristic is known exactly.

Most of the available formulae for f.m. distortion do not give correctly the effects of one or both of constant and linear parts of the phase characteristic. Carson and Fry[5] suggest that the linear portion of the phase characteristic should be subtracted before f.m. distortion is evaluated, although they do not discuss the error involved if this is not done. Gladwin,[21] using an expansion developed in an earlier paper, evaluated for a particular network the distortion with and without subtraction of the linear phase component, and decided that it was preferable to perform the initial subtraction.

Quasi-stationary theory gives the correct result for a constant phase shift, but for a linear phase shift it produces expressions for the fundamental amplitude and phase change which, although correct in the limit of $\beta = 0$, approach the limiting forms in a manner different from that of the true expressions. Taking the linear phase characteristic in the form $\phi = \tau(\omega - \omega_c)$ and the undistorted wave as $\sin(\omega_c t + \beta \sin \omega_a t)$, the distorted wave is (without approximation) $\sin\{\omega_c(t + \tau) + \beta \sin [\omega_a(t + \tau)]\}$ so that the fundamental phase change is $\omega_a \tau$ and the fundamental amplitude change is unity. In the quasi-stationary approximation, the phase modulation becomes $\beta \sin \omega_a t + \tau\beta\omega_a \cos \omega_a t$. Thus, to this approximation, the fundamental phase change is arc tan $(\omega_a t)$ and the fundamental amplitude change is $\sqrt{(1 + \omega_a^2 \tau^2)}$.

The approximate first-order theory gives for both constant and linear phase characteristics the same result as the quasi-stationary theory (see Section 9.2 of Reference 8).

The extended first-order theory predicts a phase shift of the incoming wave of magnitude $\sin \psi$ due to a constant phase characteristic of the form $\phi(\omega - \omega_c) = \psi$ (Section 10.4). While this is incorrect, no spurious distortion components have appeared. The linear phase characteristic, however, produces according to this theory not only an incorrect fundamental (like the quasi-stationary and approximate first-order theories) but also spurious harmonics (Section 10.4). The foregoing refers to characteristics which are purely constant or purely linear. When either constant or linear terms are added to more complicated

phase characteristics, the computed fundamental and harmonic expressions become more or less inaccurate.

As mentioned earlier, this unfortunate state of affairs is generally found in the various approximate formulae proposed for the regions of parameter values in frequency-modulation theory which cannot be covered by direct Fourier analysis. Approximate formulae, such as expression (6), must be used, since otherwise many frequency-modulation problems become intractable, but they must clearly be used with caution. Where a phase characteristic does not depart greatly either from linearity or constancy, experience shows that most accurate results will be obtained if the appropriate linear or constant terms are subtracted before the analysis. As an example, Fig. 5

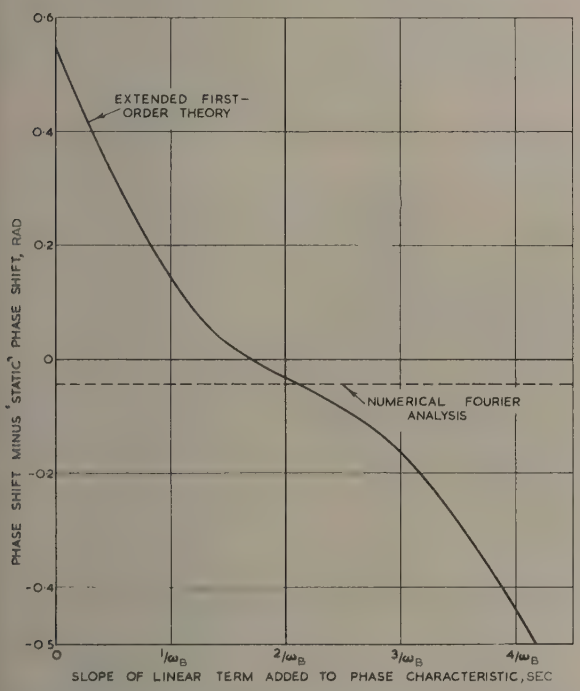


Fig. 5.—Relative phase response of the fundamental for the maximally-flat-amplitude triple-tuned circuit: effect on the extended first-order theory of the addition of a linear phase term ($q = 1$, $p = 0{\cdot}5$).
Slope of linear term of phase characteristic $= -2/\omega_B$.

shows the effect on the fundamental phase response, according to the extended first-order theory, of the addition of various amounts of linear term to the phase characteristic of the network considered in Section 4. The normalized modulating frequency is taken as 0·5, this being a value for which a first-order theory should hold reasonably well (see Figs. 1 and 2). This example shows clearly the penalty for failing to remove initially the linear part of the phase characteristic.

(8) CONCLUSIONS

When a wave frequency-modulated by a single tone traverses a passive network whose amplitude and phase characteristics are non-linear with frequency, the modulating tone undergoes amplitude and phase distortion and harmonics are generated in the modulation. When the modulation index is not too large, only a small number of significant sidebands are present. The modulation distortion can then be computed by a direct Fourier analysis. For a large modulation index the modulation distortion can be deduced from a quasi-stationary argument when the network characteristics are known analytically. The intermediate region may be dealt with by using the formula developed in the present paper. This formula, which is an extension of the approximate first-order formula derived in Reference 8, may be applied to characteristics known analytically or to measured characteristics. The regions in which the three lines of approach overlap must be determined by trial in the particular situation. From consideration of specific cases it appears that, before applying either the quasi-stationary analysis or the extended first-order theory, the linear part of the phase characteristic (defined in terms of the first derivative at carrier frequency) should be subtracted from it.

(9) REFERENCES

(1) CORRINGTON, M. S.: 'Variation of Bandwidth with Modulation Index in Frequency Modulation', *Proceedings of the Institute of Radio Engineers*, 1947, **35**, p. 1013.

(2) GLADWIN, A. S.: Discussion on 'Generalized Theory of Multitone Amplitude and Frequency Modulation', *ibid.*, 1948, **36**, p. 240.

(3) SKINNER, L. V.: 'Spectra of Waves with Periodic Modulation', *Transactions of the Institute of Radio Engineers*, June, 1954, **CT–1**, p. 26.

(4) WATT, A. D.: 'Spectrum of Frequency-Shift Radio Photo-Transmissions', *ibid.*, October, 1956, **CS–4**, p. 27.

(5) CARSON, J. R., and FRY, T. C.: 'Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency-Modulation', *Bell System Technical Journal*, 1937, **16**, p. 513.

(6) WOJNAR, A.: 'Experimental Investigation of F.M. Signal Distortion', *Prace Przemysłowym Instytutem Telekomunikacji*, 1954, **5**, p. 18 (in Polish).

(7) BORODICH, S. V.: 'Linear Distortions in Frequency-Modulation', *Radiotekhnika*, 1954, **9**, No. 6, p. 66.

(8) MEDHURST, R. G.: 'Harmonic Distortion of Frequency-Modulated Waves by Linear Networks', *Proceedings I.E.E.*, Paper No. 1650, May, 1954 (**101**, Part III, p. 171).

(9) SKWIRZYNSKI, J. K.: 'The Linear Distortion of FM Signals in Band-Pass Filters for Large Modulation Frequencies', *Marconi Review*, 1954 (4th Quarter), **17**, p. 101.

(10) COLLINGS, R. H. P., and SKWIRZYNSKI, J. K.: 'The Distortion of F.M. Signals in Passive Networks', *ibid.*, p. 113.

(11) MEDHURST, R. G., and SMALL, G. F.: 'Distortion in Frequency-Modulation Systems Due to Small Sinusoidal Variations of Transmission Characteristics', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 1608.

(12) MEDHURST, R. G., and HYAMSON, H. D.: 'Discriminator Distortion in Frequency-Modulation Systems', *Proceedings I.E.E.*, Monograph No. 226 R, February, 1957 (**104** C, p. 357).

(13) BROWN, R. F.: 'Frequency-Modulation Distortion in Linear Networks', *ibid.*, Paper No. 219 R, January, 1957 (**104** B, p. 52).

(14) BIOT, M. A.: 'A New Approach to the Non-Linear Problems of F.M. Circuits', *Quarterly Journal of Applied Mathematics*, 1957, **15**, p. 1.

(15) RICE, S. O.: 'Distortion in a Noise-Modulated F.M. Signal by Non-Linear Attenuation', *Bell System Technical Journal*, 1957, **36**, p. 879.

(16) BAGHDADY, E. G.: 'Theory of Low-Distortion Reproduction of FM Signals in Linear Systems', *Transactions of*

the Institute of Radio Engineers, September, 1958, **CT–5**, p. 202.
(17) C.C.I.R. Documents of the VIIIth Plenary Assembly, Warsaw, 1956, Volume 1, Recommendations, Reports, Resolutions, Questions and Study Programmes, International Telecommunication Union, Geneva, 1957, pp. 427 and 428.
(18) VALLEY, G. E., and WALLMAN, H.: 'Vacuum Tube Amplifiers' [McGraw-Hill (M.I.T. Radiation Laboratory Series No. 18), 1948].
(19) BECK, F.: 'Harmonic Analysis Using a Digital Computer', *Computer Journal*, 1958, **1**, p. 117.
(20) WATT, J. M.: 'A Note on the Evaluation of Trigonometric Series', *ibid.*, 1959, **1**, p. 162.
(21) STUMPERS, F. L. H. M., and GLADWIN, A. S.: Discussion on 'The Distortion of Frequency-Modulated Waves by Transmission Networks', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1257.
(22) DWIGHT, H. B.: 'Tables of Integrals and Other Mathematical Data' (Macmillan, New York, 1947), Revised Edition.
(23) WHITTAKER, E. T., and WATSON, G. N.: 'A Course of Modern Analysis' (Cambridge University Press, 1952), Fourth Edition.
(24) TERMAN, F. E.: 'Radio Engineers' Handbook' (McGraw-Hill, 1943).
(25) MCLACHLAN, N. W.: 'Bessel Functions for Engineers' (Oxford University Press, 1955), Second Edition.

(10) APPENDICES

(10.1) Derivation of the Extended First-Order Formula

Taking the undistorted wave as $\sin(\omega_c t + \beta \sin \omega_a t)$, the corresponding spectrum is

$$\sum_{n=-\infty}^{\infty} J_n \sin[(\omega_c + n\omega_a)t]$$

After passage through the distorting network, this becomes

$$\sum_{n=-\infty}^{\infty} J_n[1 + \rho(\omega_a + n\omega_a)] \sin[(\omega_c + n\omega_a)t + \phi(\omega_c + n\omega_a)]$$

From this expression we pick out Δ_1 and Δ_2. It is found after some rearrangement that

$$\Delta_1 = \sum_{n=-\infty}^{\infty} J_n\Big(\{[1 + \rho(\omega_c + n\omega_a)] \cos[\phi(\omega_c + n\omega_a)] - 1\} \cos(n\omega_a t) - [1 + \rho(\omega_c + n\omega_a)] \sin[\phi(\omega_c + n\omega_a)] \sin(n\omega_a t)\Big) \quad (21)$$

and

$$\Delta_2 = \sum_{n=-\infty}^{\infty} J_n\Big(\{[1 + \rho(\omega_c + n\omega_a)] \cos[\phi(\omega_c + n\omega_a)] - 1\} \sin(n\omega_a t) + [1 + \rho(\omega_c + n\omega_a)] \sin[\phi(\omega_c + n\omega_a)] \cos(n\omega_a t)\Big) \quad (22)$$

These expressions have now to be inserted in eqn. (4), and the various harmonic distortion terms extracted. After some algebraic manipulation one arrives at the Nth harmonic component in the form given as expression (6).

(10.2) Quasi-Stationary Approximation for Minimum-Phase Network having Linear Amplitude Slope

In order to find the coefficient of the harmonic terms in expression (17) it is necessary to write

$$\log\left|1 - \frac{\beta\omega_a}{\omega_\Delta} \cos \omega_a t\right|$$

in harmonic form. The nth harmonic coefficient of this expression is given by

$$\frac{2}{\pi}\int_0^\pi \log|1 - K\cos\theta| \cos n\theta d\theta \quad . \quad . \quad . \quad (23)$$

where K has been written for $\beta\omega_a/\omega_\Delta$, and θ for $\omega_a t$. Integrating expression (23) by parts, we obtain

$$\left[\frac{2}{\pi}\frac{1}{n}\sin n\theta \log|1 - K\cos\theta|\right]_0^\pi - \frac{2}{n\pi}\int_0^\pi \frac{\sin n\theta\, K \sin\theta}{1 - K\cos\theta} d\theta = \frac{K}{n\pi}\int_0^\pi \frac{\cos[(n-1)\theta] - \cos[(n+1)\theta]}{1 - K\cos\theta} d\theta \quad (24)$$

Thus, from expression (17), the second-harmonic term in the phase modulation is given by

$$\frac{\beta\omega_a k}{8\cdot 686}\left(-\frac{1}{2\pi}\int_0^\pi \frac{\cos\theta - \cos 3\theta}{1 - K\cos\theta} d\theta + \frac{K}{2\pi}\int_0^\pi \frac{1 - \cos 2\theta}{1 - K\cos\theta} d\theta + \frac{K}{6\pi}\int_0^\pi \frac{\cos 2\theta - \cos 4\theta}{1 - K\cos\theta} d\theta\right) = \frac{\beta\omega_a k}{8\cdot 686 \times 6\pi}\int_0^\pi \frac{3K - 3\cos\theta - 2K\cos 2\theta + 3\cos 3\theta - K\cos 4\theta}{1 - K\cos\theta} d\theta$$

which, after some manipulation, reduces to

$$\frac{\beta\omega_a k}{8\cdot 686}\left[\frac{3K^2 - 2}{3K^3} + \frac{2}{3\pi K^3}(K^2 - 1)^2 \int_0^\pi \frac{d\theta}{1 - K\cos\theta}\right] \quad (25)$$

When $K < 1$, the integral is known (Reference 22, integral 859·21), and expression (25) reduces to

$$\frac{\beta\omega_a k}{8\cdot 686}\left[\frac{3K^2 - 2}{3K^3} + \frac{2(1 - K^2)^{3/2}}{3K^3}\right] \quad . \quad . \quad . \quad (26)$$

When $K > 1$ we evidently require the Cauchy principle value of the integral (Reference 23), p. 75. It is thus required to find

$$\lim_{\delta_1 \to \delta_2 \to 0}\left[\int_0^{\arccos(1/K) - \delta_1} \frac{d\theta}{1 - K\cos\theta} + \int_{\arccos(1/K) + \delta_2}^{\pi} \frac{d\theta}{1 - K\cos\theta}\right]$$

From Reference 22, integral 446·00, this is

$$\lim_{\delta_1 \to \delta_2 \to 0}\left\{\left[\frac{2}{\sqrt{(K^2 - 1)}} \operatorname{arc\,tanh} \frac{(-K - 1)\tan\frac{\theta}{2}}{\sqrt{(K^2 - 1)}}\right]_0^{\arccos(1/K) - \delta_1} + \left[\frac{2}{\sqrt{(K^2 - 1)}} \operatorname{arc\,coth} \frac{(-K - 1)\tan\frac{\theta}{2}}{\sqrt{(K^2 - 1)}}\right]_{\arccos(1/K) + \delta_2}^{\pi}\right\}$$

In the limit of small values of δ_1 and δ_2 the expression in the curly brackets approaches

$$\frac{1}{\sqrt{(K^2-1)}}\log\frac{\delta_1}{\delta_2}$$

Thus, letting δ_1 and δ_2 approach zero at the same rate, the Cauchy principle value of the integral becomes zero. The second-harmonic coefficient in the phase modulation when $K>1$ is thus

$$\frac{\beta\omega_a k}{8\cdot 686}\left(\frac{3K^2-2}{3K^3}\right) \quad . \quad . \quad . \quad . \quad . \quad (27)$$

From expressions (26) and (27), after differentiating the phase modulation with respect to time to recover the frequency modulation, we arrive at a second-harmonic ratio given by expression (18).

(10.3) **Proof that the Limiting Form of the Approximate First-Order Expression is the Quasi-Stationary Approximation**

From expression (5), the Nth harmonic term due to the phase characteristic, according to the approximate first-order theory, is

$$\cos(N\omega_a t)\sum_{n=-\infty}^{\infty} J_n\left(\frac{\omega_d}{\omega_a}\right)\left[J_{n-N}\left(\frac{\omega_d}{\omega_a}\right)+J_{n+N}\left(\frac{\omega_d}{\omega_a}\right)\right]\phi(\omega_c+n\omega_a)$$

$$=\cos(N\omega_a t)\sum_{n=0}^{\infty}{}' J_n\left(\frac{\omega_d}{\omega_a}\right)\left[J_{n-N}\left(\frac{\omega_d}{\omega_a}\right)+J_{n+N}\left(\frac{\omega_d}{\omega_a}\right)\right]$$

$$[\phi(\omega_c+n\omega_a)+(-1)^N\phi(\omega_c-n\omega_a)] \quad . \quad . \quad (28)$$

where the argument of the Bessel functions has been written out explicitly, and Σ' denotes that the term in $(-1)^N$ is omitted for $n=0$. It will now be shown that, in the limit of small values of ω_a, the coefficient of the Nth harmonic, written as an infinite series in expression (28), can be represented adequately by the truncated series.

$$\sum_{n=0}^{\omega_d/\omega_a}{}' J_n\left(\frac{\omega_d}{\omega_a}\right)\left[J_{n-N}\left(\frac{\omega_d}{\omega_a}\right)+J_{n+N}\left(\frac{\omega_d}{\omega_a}\right)\right]$$

$$[\phi(\omega_c+n\omega_a)+(-1)^N\phi(\omega_c-n\omega_a)] \quad (29)$$

This is associated with the result, well known in frequency-modulation theory, that as ω_a decreases relative to ω_d the sideband energy is increasingly confined to the frequency region between $\omega_c-\omega_d$ and $\omega_c+\omega_d$ [see, for example, Fig. 50(*b*) of Reference 24]. The proof requires rather careful use of limiting formulae for the Bessel functions. What will actually be proved is that

$$\lim_{\omega_a\to 0}\frac{J_{\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\left[J_{-N+\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)+J_{N+\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\right]}{J_{\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\left[J_{-N+\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)+J_{N+\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\right]}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

for any fixed values of δ, however small.

So long as δ is not zero, the terms of the numerator of the left-hand-side of eqn. (30) can be expanded using expression (1), paragraph 1, Section 4.35 of Reference 25. It is found that

$$J_{N+\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{\exp\left\{-\frac{1}{3}\frac{\omega_d}{\omega_a}\left(1+\delta+\frac{N\omega_a}{\omega_d}\right)\left[\left(1+\delta+\frac{N\omega_a}{\omega_d}\right)^2-1\right]^{3/2}\right\}}{\sqrt{(2\pi)}\left(\frac{\omega_d}{\omega_a}\right)^{1/2}\left[\left(1+\delta+\frac{N\omega_a}{\omega_d}\right)^2-1\right]^{1/4}} \quad . \quad . \quad . \quad . \quad . \quad (31)$$

$$J_{-N+\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{\exp\left\{-\frac{1}{3}\frac{\omega_d}{\omega_a}\left(1+\delta-\frac{N\omega_a}{\omega_d}\right)\left[\left(1+\delta-\frac{N\omega_a}{\omega_d}\right)^2-1\right]^{3/2}\right\}}{\sqrt{(2\pi)}\left(\frac{\omega_d}{\omega_a}\right)^{1/2}\left[\left(1+\delta-\frac{N\omega_a}{\omega_d}\right)^2-1\right]^{1/4}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

$$J_{\omega_d(1+\delta)/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{\exp\left\{-\frac{1}{3}\frac{\omega_d}{\omega_a}(1+\delta)[(1+\delta)^2-1]^{3/2}\right\}}{\sqrt{(2\pi)}\left(\frac{\omega_d}{\omega_a}\right)^{1/2}[(1+\delta)^2-1]^{1/4}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

The terms in the denominator can be expanded using expression (1), paragraph 2, Section 4.35 of Reference 25. It is found that

$$J_{N+\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{1}{48^{1/6}}\frac{\Gamma(1/3)}{\pi\left(\frac{\omega_d}{\omega_a}+N\right)^{1/3}} \quad . \quad . \quad (34)$$

$$J_{-N+\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{1}{48^{1/6}}\frac{\Gamma(1/3)}{\pi\left(\frac{\omega_d}{\omega_a}-N\right)^{1/3}} \quad . \quad . \quad (35)$$

$$J_{\omega_d/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\to\frac{1}{48^{1/6}}\frac{\Gamma(1/3)}{\pi\left(\frac{\omega_d}{\omega_a}\right)^{1/3}} \quad . \quad . \quad . \quad (36)$$

Substituting for the Bessel functions in the left-hand-side of eqn. (30), using relations (31)–(36), and letting ω_a approach zero while ω_d and δ are fixed, it is apparent that eqn. (30) is correct in the limit.

Returning to expression (29), put $n=z/\omega_a$, so that the summation becomes

$$\sum_{z=0}^{\omega_d}{}' J_{z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\left[J_{-N+z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)+J_{N+z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\right][\phi(\omega_c+z)+(-1)^N\phi(\omega_c-z)]$$

where z increases in steps of width ω_a. As ω_a approaches zero, the summation tends to the integral form

$$\int_0^{\omega_d} J_{z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\left[J_{-N+z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right) + J_{N+z/\omega_a}\left(\frac{\omega_d}{\omega_a}\right)\right] [\phi(\omega_c + z) + (-1)^N\phi(\omega_c - z)]\frac{dz}{\omega_a}$$

The Bessel functions can be replaced by the first term of their asymptotic expansions [Reference 25, Section 4.30, expression (1)] over the whole of the range of integration without serious error. After some rearrangement, the integral becomes

$$\int_0^{\omega_d} \frac{2}{\pi}\frac{1}{\sqrt{\left[\left(\frac{\omega_d}{\omega_a}\right)^2 - \left(\frac{z}{\omega_a}\right)^2\right]}}\left(\cos\left\{2\sqrt{\left[\left(\frac{\omega_d}{\omega_a}\right)^2 - \left(\frac{z}{\omega_a}\right)^2\right]} - \frac{2z}{\omega_a}\arccos\left(\frac{z}{\omega_d}\right) - \frac{\pi}{2}\right\} + 1\right)\cos\left(N\arccos\frac{z}{\omega_d}\right) [\phi(\omega_c + z) + (-1)^N\phi(\omega_c - z)]\frac{dz}{\omega_a}$$

As ω_a decreases, the contribution to the integral from the first cosine term will decrease, owing to cancellation between successive closely-spaced cycles. Thus the integral will approach the form

$$\int_0^{\omega_d} \frac{2}{\pi}\frac{1}{\sqrt{(\omega_d^2 - z^2)}}\cos\left(N\arccos\frac{z}{\omega_d}\right) [\phi(\omega_c + z) + (-1)^N\phi(\omega_c - z)]dz \quad (37)$$

Now put $z = \omega_d \cos \omega_a t$, so that expression (37) becomes

$$\frac{2}{\pi}\int_0^{\pi/2} \cos(N\omega_a t)[\phi(\omega_c + \omega_d\cos\omega_a t) + (-1)^N\phi(\omega_c - \omega_d\cos\omega_a t)]d(\omega_a t) \quad (38)$$

The final transformation required is to write $\pi - \omega_a t$ for $\omega_a t$ in the second ϕ term. Then expression (38) becomes

$$\frac{2}{\pi}\int_0^{\pi} \phi(\omega_c + \omega_d\cos\omega_a t)\cos(N\omega_a t)d(\omega_a t)$$

and this will be recognized as the quasi-stationary formula for the Nth harmonic.

(10.4) Effect of Constant and Linear Phase Characteristics according to the Extended First-Order Theory

(10.4.1) Constant Phase Characteristic.

Let $\phi(\omega_c + n\omega_a) = \psi$, where ψ is independent of n, and take $\rho(\omega_c + n\omega_a) = 0$.

From expression (21)

$$\begin{aligned}\Delta_1 &= \sum_{n=-\infty}^{\infty} J_n[(\cos\psi - 1)\cos n\omega_a t - \sin\psi\sin n\omega_a t] \\ &= (\cos\psi - 1)\cos(\beta\sin\omega_a t) - \sin\psi\sin(\beta\sin\omega_a t) \\ &= \cos(\beta\sin\omega_a t + \psi) - \cos(\beta\sin\omega_a t) \\ &= \cos(\mu_t + \psi) - \cos\mu_t\end{aligned}$$

Similarly, from expression (22)

$$\Delta_2 = \sin(\mu_t + \psi) - \sin\mu_t$$

Then, from expression (4),

$$\begin{aligned}\Delta &= \Delta_2\cos\mu_t - \Delta_1\sin\mu_t \\ &= \sin\psi\end{aligned}$$

(10.4.2) Linear Phase Characteristic.

Let $\phi(\omega_c + n\omega_a) = \tau n\omega_a$, and $\rho(\omega_c + n\omega_a) = 0$. From expression (21),

$$\begin{aligned}\Delta_1 &= \sum_{n=-\infty}^{\infty} J_n[(\cos\tau n\omega_a - 1)\cos n\omega_a t - \sin\tau n\omega_a\sin n\omega_a t] \\ &= 2\sum_{n=1}^{\infty} J_{2n}\cos[2n\omega_a(t+\tau)] - 2\sum_{n=1}^{\infty} J_{2n}\cos 2n\omega_a t \\ &= \cos\{\beta\sin[\omega_a(t+\tau)]\} - \cos(\beta\sin\omega_a t)\end{aligned}$$

and similarly $\Delta_2 = \sin\{\beta\sin[\omega_a(t+\tau)]\} - \sin(\beta\sin\omega_a t)$

Then, from expression (4),

$$\begin{aligned}\Delta &= \Delta_2\cos(\beta\sin\omega_a t) - \Delta_1\sin(\beta\sin\omega_a t) \\ &= \sin[2\beta\sin\tfrac{1}{2}\omega_a\tau\cos(\omega_a t + \tfrac{1}{2}\omega_a\tau)] \\ &= 2J_1\cos(\omega_a t + \tfrac{1}{2}\omega_a\tau) - 2J_3\cos 3(\omega_a t + \tfrac{1}{2}\omega_a\tau)\ldots\end{aligned}$$

where the argument of the Bessel functions is $2\beta\sin\frac{1}{2}\omega_a\tau$.

621.385.3: 621.395.64

The Institution of Electrical Engineers
Paper No. 3163 E
Mar. 1960
©

POST OFFICE VALVES FOR DEEP-WATER SUBMARINE TELEPHONE REPEATERS

By M. F. HOLMES, .Sc., and F. H. REYNOLDS, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 21*st May, and in revised form* 21*st September*, 1959.)

SUMMARY

As the length of submarine telephone-cable systems is increased, new demands are placed on the amplifying valves used in the submerged repeaters. Based on the experience of earlier valves suitable for shallow-water repeaters, two new valves have been developed for deep-water submarine cables.

(1) INTRODUCTION

The historical development of the thermionic valves used by the Post Office in their submarine repeaters in shallow waters round the British Isles, and in the Newfoundland–Nova Scotia section of the first transatlantic telephone cable, has been summarized in an earlier paper.[1] It was realized, however, that in the event of British transoceanic telephone cables being required, limitations of cable and equipment insulation would lead to a general lowering of overall repeater voltages which in turn would necessitate new valves. In particular, the 124 volts of the Newfoundland–Nova Scotia repeater,[2] using type 6P12 valves, was much too high for a transatlantic crossing.

The development of deep-water valves was commenced before the laying (in 1955–56) of the first transatlantic telephone cable and continued steadily until the authorization of the telephone cable between Britain and Canada lent urgency to the project.

The new valves, known as the types 10P1 and 10P2 (for output and input use, respectively) embody experience gained on the type 6P12 valve and the results of further valve research. Special attention has been paid to:

(*a*) Further improvement in the long-term stability of the principal valve parameters with a new emphasis on insurance against any deterioration in quality in the course of a production run.
(*b*) The avoidance of breakdown of the heater-cathode insulation, and the elimination of effects associated with the deposition of conducting films on the valve insulators.
(*c*) The reduction of h.t. voltage without the loss of output power.

The present paper discusses both the valves and their development problems; some operating data and an abridged performance specification are also included.

(2) LIFE STABILITY OF THE PRINCIPAL VALVE PARAMETERS

The basic gain of a repeater amplifier is determined by the mutual conductance of the valves. Moreover, because the mutual conductance is a function of the anode current which in turn depends on the working-point of the valve (in a given auto-bias circuit), stability of both the mutual conductance and the working-point are matters of primary importance. Evidently the passage of current through the control-grid resistor will also disturb the steady anode current, so that the limitation of control-grid current ranks equally with the stability of mutual conductance and working-point; all three quantities are reviewed in the following Sections.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Holmes and Mr. Reynolds are at the Post Office Research Station.

(2.1) Mutual Conductance

To achieve stability of mutual conductance, it is necessary to prevent the growth of a resistive interface between the cathode coating and core, and to maintain the activity of the cathode at an adequate level.

At an early stage[1,3] in the development of valves for submerged repeaters, platinum was substituted for silicon-bearing nickel as a cathode-core material. In this way, any possibility of interface growth has been avoided for almost all the Post Office valves now operational in submarine cables. Concurrently with the production of platinum-core valves for planned and authorized systems, an experimental programme was undertaken for the purpose of assessing tungsten-nickel alloy[4] as an alternative core material. Experiments have now demonstrated that, over a period of 40 000 hours of life tests, the trend of mutual conductance for valves with cathode cores of tungsten-nickel alloy is not inferior to the corresponding trend for platinum-core valves. Two typical life-test curves are shown in Fig. 1, from which it can be seen that both materials yield valves equally free from growth of interface resistance.

If the comparison between tungsten-nickel alloy and platinum is extended to other properties, it is found that the alloy begins to offer distinct advantages. Not only is the alloy stronger mechanically but the finished cathode is easier to activate, so that after the very stringent mechanical and electrical tests at the time of manufacture, typical shrinkage rates may be about 30% for tungsten-nickel alloy but as much as 60% for valves with platinum cores. A specification of the tungsten-nickel alloy (coded L2) used by the Post Office is given in Section 9.1.

The level of activity of the oxide-coated cathode exerts a direct influence on the mutual conductance through the operation of the Langmuir space-charge current/voltage relationship for a thermionic diode. There is also evidence[5] that the ohmic resistance of the cathode matrix is inversely proportional to the emission level, subject to an upper resistance limit[6] (of the order of 100 ohms in Post Office valves). Because any growth of matrix resistance would depress the mutual conductance in exactly the same way as does the development of interface resistance, it is doubly important to maintain an adequate level of cathode emission.

The maintenance of an adequate emission level has been resolved into one of avoiding deactivation of the cathode by gas attack. It is not proposed to consider here any of the techniques which have been evolved in the attempt to solve this problem except in one instance when the elimination of a source of gas resulted in a change of material in the valve.

It has been demonstrated that if soda glass is used for the valve envelope, a danger of glass electrolysis exists which may lead to the evolution of a gas—probably oxygen—into the valve atmosphere. This can be avoided either by taking special precautions to eliminate electrolysing voltages across the glass or, in a simpler manner, by using lead glass which has a much higher resistivity.

An estimate of the degree of success so far achieved in main-

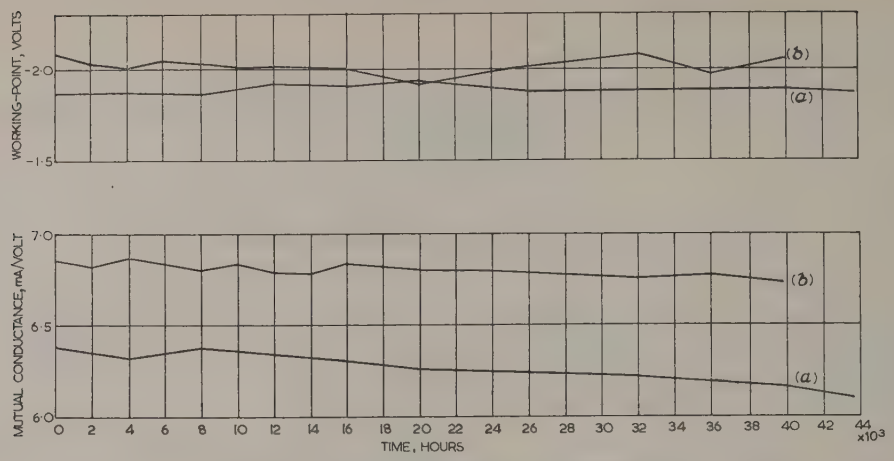


Fig. 1.—Life stability of mutual conductance and working-point in valves.
(*a*) Platinum cathode core.
(*b*) Tungsten-nickel-alloy cathode core.

taining a stable level of mutual conductance can be assessed in terms of curves such as those of Fig. 1.

(2.2) Working-Point

The working-point of a valve (defined as that voltage externally applied to the control grid to yield a given anode current at specified screen-grid and anode potentials) is a function of (*a*) electrode geometry, (*b*) ohmic resistance in the cathode and (*c*) contact potential between the control grid and cathode. Of these factors, only cathode resistance and contact potential are important for stability, and if the cathode ohmic resistance has been minimized sufficiently for the promotion of mutual conductance stability, the invariance of the working-point on this account is automatically assured.

It is well known that the stability of the control-grid/cathode contact potential is delicately dependent on the ageing schedules used at the time of manufacture. If, however, valves with cathode cores of tungsten-nickel alloy are subjected to a 'long-ageing' process amounting to about 2000 hours of operation under normal working conditions, the contact potential reaches a value (of approximately 0·7 volt) which is subsequently very stable and moreover varies very little from valve to valve. Typical life stability of working-point in such valves is also illustrated by Fig. 1.

(2.3) Grid Current

The development of control-grid current during life may be caused by (*a*) grid emission, (*b*) positive ion current or (*c*) leakage.

It is common practice to inhibit grid emission by coating the grid wires with a layer of gold,[7] but this process imposes severe restrictions on the temperature to which the grid can be raised whilst the valve is being pumped and aged. If the control grid is designed so that its operating temperature does not exceed 240° C, then, with a cathode core of tungsten-nickel alloy, gold coating is unnecessary.

From the steps successfully taken to mitigate gas deactivation of the cathode in the interest of stabilizing the mutual conductance, the positive-ion component of the control-grid current can be reduced below the X-ray level.[8]

The development of control-grid current due to leakage is only one of the several manifestations of the deposition of conducting films on the insulating surfaces of valves, so that the topic will be held over until the next Section. It will suffice to say at the present juncture that, by proper attention to the causes of grid current, stable values of less than 0·01 μA can be maintained for at least 4 years.

(3) SECONDARY FORMS OF ELECTRICAL FAILURE

Thermionic valves made to the formula outlined in the previous Section are likely to be free of the commonest forms of valve failure. They will, however, still be liable to secondary defects which, after sufficient time, may seriously jeopardize the proper functioning of a repeater amplifier. Some of these defects are associated with the deposition of conducting films on the insulating surfaces of the valve. Another problem arises from the method of economizing on repeater voltage by disposing the valve heaters in series so that each carries the direct line current, the h.t. voltage then being derived from the potential drop across the heater chain. The valve heaters are then, in general, at a positive potential relative to the cathode sleeves.

(3.1) Growth of Inter-Electrode Capacitances and Noise

In addition to the onset of leakage, the deposition of conducting films on mica insulators can cause a marked increase in inter-electrode capacitances.[9] The anode-to-grid capacitance is particularly vulnerable and may increase several-fold after one or two thousand hours' operation. Such valves are also likely to generate random bursts of noise of relatively high intensity.[10]

It has long been known that the oxide cathode is the source of unwanted conducting films deposited in valves.[11] When the platinum cathode was first introduced, however, a substantial diminution in barium evaporation was observed.[12] It seems probable that cathodes having cores of tungsten-nickel alloy possess evaporation properties characteristic of platinum-core valves. Recently, other low-order evaporation effects have been observed, originating at the anode,[9] which are nevertheless of serious magnitude when life spans exceeding several thousand hours are required. By the use of very pure anode materials—tungsten-nickel (L2) alloy is again suitable—the anode can be eliminated as a source of evaporation. If, furthermore, arrangements are made to shield vital areas of the insulating surfaces from direct evaporation, film effects can be reduced to a very low order.

(3.2) Heater-Cathode Insulation

If a conventional sprayed and sintered heater is operated at voltage of some tens of volts (or more) positive with respect o its enveloping cathode sleeve, a breakdown effect is likely to ccur, after sufficient time, in which the insulation resistance alls to the order of 10^4–10^5 ohms. This behaviour has been escribed elsewhere,[13] and it has also been demonstrated that he effect can be avoided by rendering the alumina non-porous. The practical solution to the problem lies in the use of a rigid nsulator made of non-porous pure alumina which, from the esults of accelerated breakdown tests, promises an indefinite fe under the conditions in a submerged repeater.

(4) PROBLEMS OF VALVE DESIGN IN THE FACE OF RESTRICTED H.T. VOLTAGE

The general problem of designing a pentode valve to satisfy a arget specification has been discussed elsewhere.[14] For sub-nerged repeater valves, the interest centres on the h.t. restrictions vhich the valve must tolerate.

(4.1) Triode Section

The triode section of any receiving-type pentode valve is ormally required to satisfy a specification of anode current, vorking-point and mutual conductance, with given limitations n the maximum screen-grid voltage and current. Now, if the lectrode voltages and currents for the triode section are specified, here always exists an optimum disposition of the grid wires vhich yields the highest mutual conductance. The attainment f optimum conditions requires a balance to be achieved; on he one hand the inter-electrode gaps should be as small as ossible, because decreasing the cathode spacing from the ontrol-grid increases the mutual conductance (from the funda-nental properties of a space-charge-limited diode), and closing he gap between the control and screen grids allows the use of a ontrol grid of smaller pitch, which again raises the mutual onductance. On the other hand, the gaps must not be too nall, otherwise local field variations due to the proximity of he grid wires to neighbouring electrodes will bring about a eduction in the mutual conductance. When the above optima ave been achieved, however, the design will still be subject to verriding limitations which for any accepted method of valve onstruction can be expressed by the symbolic equation

$$\begin{bmatrix}\text{potential}\\ \text{of screen}\\ \text{grid}\end{bmatrix} + \begin{bmatrix}\dfrac{1}{\text{mechanical}}\\ \text{quality}\end{bmatrix} = \begin{bmatrix}\text{mutual}\\ \text{conduc-}\\ \text{tance}\end{bmatrix} + \begin{bmatrix}\text{magnitude}\\ \text{of working}\\ \text{point}\end{bmatrix} + \begin{bmatrix}\text{steady}\\ \text{anode}\\ \text{current}\end{bmatrix}$$

The measure of mechanical quality in this context is partly he size of the inter-electrode gaps, particularly between control rid and cathode, and partly the diameter of the grid wires, articularly the control-grid wire. It will be clear from the bove equation that an increase in any of the quantities on he right-hand side can be obtained only at the expense of at east one of the other two, unless it is permissible either to ncrease the potential of the screen grid or degrade the mechanical uality of the valve.

Looked at another way, the screen-grid voltage is in the ature of a credit which can be distributed as desired between lectrical performance and mechanical quality. For the sub-nerged repeater valve, this credit is very restricted so that the ttainment of the optimum conditions described above becomes ighly desirable. In the apportionment of this credit it should be realized that probably no valve enjoys a more undisturbed life, mechanically, than does a repeater valve lying on the ocean bed, so that no particularly high standard of structural rugged-ness is necessary. Ruggedness, however, should not be confused with reliability, which must of of the very highest order.

(4.2) Suppressor Section

For the maximum undistorted power output from any receiving-type output valve, the position of the knee in that anode characteristic corresponding to the maximum anode-current swing should be sharply defined and located as near zero anode voltage as possible. At the same time, of course, secondary emission must be adequately suppressed, the anode impedance made as high as possible and the anode-to-grid capacitance minimized. In designing the suppressor section of an output-stage submerged-repeater valve, the satisfactory recon-ciliation of these claims is aggravated by two effects.

First, because the maximum anode-voltage swing is restricted, it is desirable to step up the steady anode current in order to obtain a useful power output. The combination of low electron velocities (due to the restricted electrode voltages) and high space currents yields high space-charge concentrations in the suppressor section. As a result, a marked potential depression readily forms in the suppressor section and, irrespective of the effect of the suppressor grid itself, the valve tends to be inherently over-suppressed.

Secondly, in any receiving-type pentode valve, the definition of the knee in the anode characteristics worsens as the screen-grid voltage increases. This is because the screen grid and anode are, as it were, competing for the space current. Usually it can be arranged for the screen-grid potential to be less than that of the anode. For the submerged repeater valve, however, it is essential to use all the available potential for the screen grid in order to obtain the best performance from the triode section. As a result, the designer has to accept screen-grid and anode voltages which are approximately equal.

The problem of obtaining a sharply defined knee in these adverse circumstances hinges on the provision of just enough secondary electron suppression and no more. On account of the inherent tendency to over-suppression, the suppressor grid will usually be of very open mesh. It would in fact seem possible to omit the suppressor-grid altogether and merely enlarge the spacing between screen-grid and anode until optimum suppression is achieved, but this often produces an intolerable drop in anode impedance and rise in anode-to-grid capacitance.

(5) DESIGN AND PERFORMANCE OF TYPE 10P VALVES

The type 10P valves have been designed to satisfy the electrical specification, shown (abridged) in Section 9.2, at the same time drawing extensively on the experience gained from the type 6P12 valve and incorporating particularly the principles and results set out in Sections 2–4.

(5.1) Heater and Cathode Structure

The heater-cathode insulation of the type 10P valves is provided by a ceramic tube, known as a heater sleeve, of flat-oval section and having a thin wall of approximately 0·15 mm thickness. The ceramic is a non-porous form of aluminium oxide of purity not less than 99·5%. The heater sleeve permits the use of a conventional lamp-coil heater dissipating 3 watts at 10 volts. The heater is wound in aluminated-tungsten wire coated with sintered pure alumina to provide limb-to-limb insulation. The diameter of the heater wire is 90 μ, which is slightly larger than that used successfully in type 6P12 valves.

The advantages of a relatively massive ceramic insulator are

not gained without some penalties. First, it will be appreciated that, for the efficient utilization of the cathode-sleeve area, the major/minor axis ratio of the cathode sleeve should be as large as possible, values of 4 being commonly used. Owing to the use of the heater sleeve, however, the cathode major/minor axis ratio in the type 10P valves is only 2·6. Secondly, it is well known that, in designing a valve heater, the lowest possible heater temperature (for a given cathode temperature) is achieved by packing as much heat-dissipating conductor into the cathode as possible. This objective is evidently jeopardized by the use of a heater sleeve, and the heater in the type 10P valve operates at 1 370° K instead of at 1 310° K as in the type 6P12 valve.

The cathode sleeve, fabricated from L2 alloy 0·075 mm thick, operates at a temperature of 1 020° K. The heater and cathode components are shown in Fig. 2.

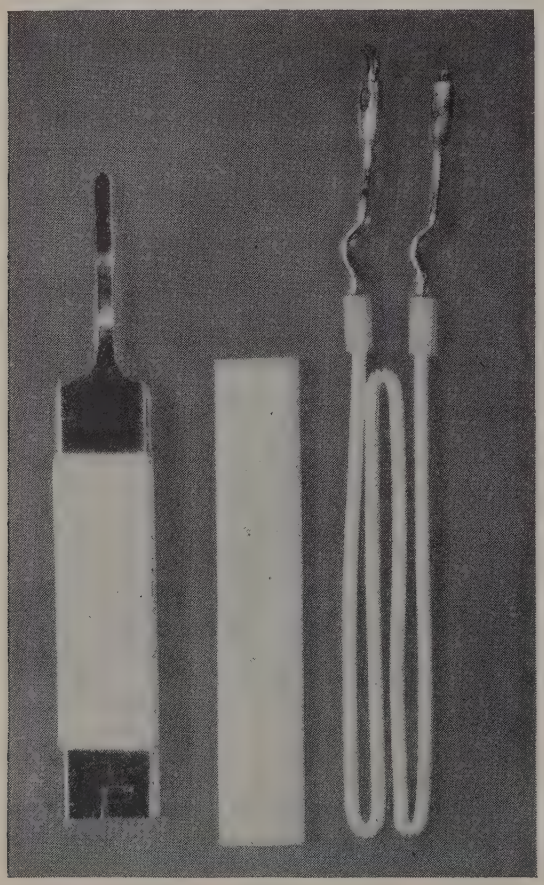


Fig. 2.—Components of the heater-cathode system of type 10P valves.

(5.2) Triode Section

The critical decision in designing the triode section is the limit to which the control-grid wire diameter and the control-grid/cathode spacing can be reduced before there exists a risk of (*a*) mechanical failure either by grid-wire fracture or by contact between a grid wire and the cathode, or (*b*) grid emission. In reaching this decision, the degree of success with which loose foreign particles can be eliminated from the interior of the envelope must be anticipated.

For the type 6P12 valve, 28 μ molybdenum wire has been used successfully, but having regard to the slightly greater span of the type 10P control grid, it was decided that a lower limit of 32 μ should be imposed for the type 10P2 valve. On the type 10P1 valve it was found possible to satisfy the electrical specification with 38 μ wire, which is an acceptable feature on account of the higher electrode dissipations compared with the type 10P2 valve.

A further less-critical decision must be made with respect to the screen-grid wire diameter, the lower limit being determined by mechanical strength. For both type 10P valves, 44 μ wire has been used.

The above dimensions having been settled, the valves were designed by the method proposed by Liebmann[15] and subsequently developed by Reynolds.[14] After some experimental adjustments for optimum performance, the typical characteristics shown in Fig. 3 were obtained. It can be seen that the margin

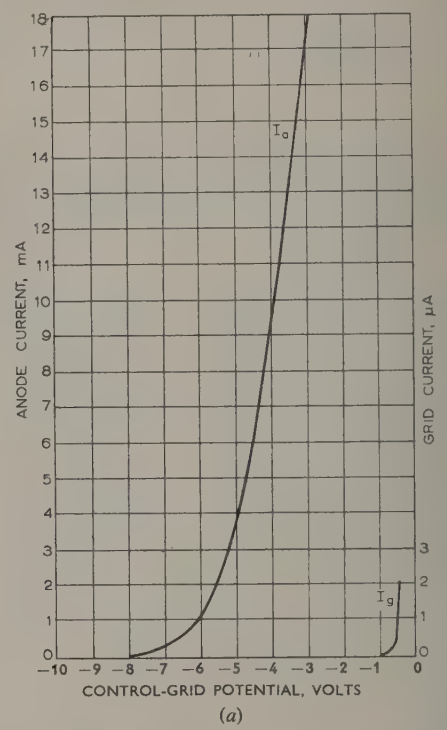


Fig. 3.—Typical mutual and grid-current characteristics.
(*a*) Type 10P1 valve. Screen-grid potential, 70 volts.
(*b*) Type 10P2 valve. Screen-grid potential, 50 volts.

from the working-point to the onset of grid current (0·1 μA) is approximately 3 volts for the type 10P1 valve and 0·5 volt for the type 10P2. The grid-to-cathode distances measured from the outermost surface of the sprayed cathode to the centre-line of the control-grid wires amount to 0·12 and 0·11 mm, respectively, for the types 10P1 and 10P2 valves. The wire centre-line spacing between the control and screen grids is 0·34 mm for both valves. The pitches of the control grids are 0·221 and 0·149 mm, respectively, for the types 10P1 and 10P2 valves, the pitch of the screen grid for both valves being 0·339 mm. In order to avoid grid emission, the control grid is fitted with graphite-coated fins which limit the control-grid-wire temperatures to a maximum of 240° C under normal operating conditions.

(5.3) Suppressor Section

Typical anode characteristics for the type 10P1 valve are shown in Fig. 4. It will be seen that the knee occurs at about 30 volts,

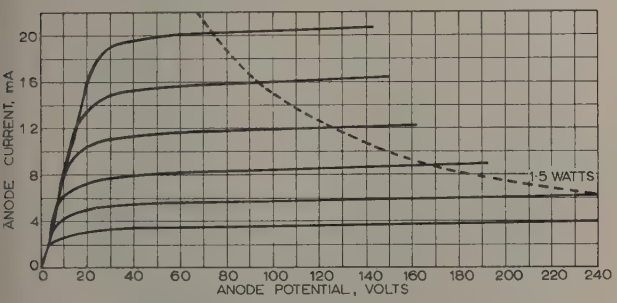


Fig. 4.—Typical anode characteristics of the type 10P1 valve.
Screen-grid potential, 70 volts.

which allows the valve to deliver up to 200 mW of signal power under conditions of optimum load. The suppressor grid of the type 10P1 valve is wound to 12·5 turns per inch.

The normal anode dissipation of the type 10P1 valve is 1·7 watt. Tests have, however, been carried out which demonstrate that the life stability of the valve is not impaired by raising the anode dissipation to 1·5 watts. This limit is shown in Fig. 4.

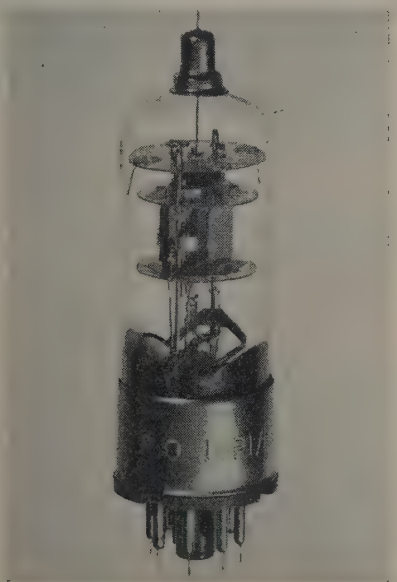


Fig. 5.—Type 10P1 valve.

(5.4) Mechanical Structure

The electrode structure is mounted on mica insulators, the inner surfaces of which are protected by means of shields standing clear of the mica, as shown in Fig. 5. The design of the shields, which are made in L2 alloy, is such that no cathode or anode evaporation can cause a film path to exist from the anode to any other electrode, nor from the control grid to any other electrode at h.t. voltage.

Retention of the heater sleeve poses a somewhat novel problem because it is not possible to notch or deform the sleeve as is common practice with metal components. In the arrangement shown in Fig. 6, the sleeve is retained at one end

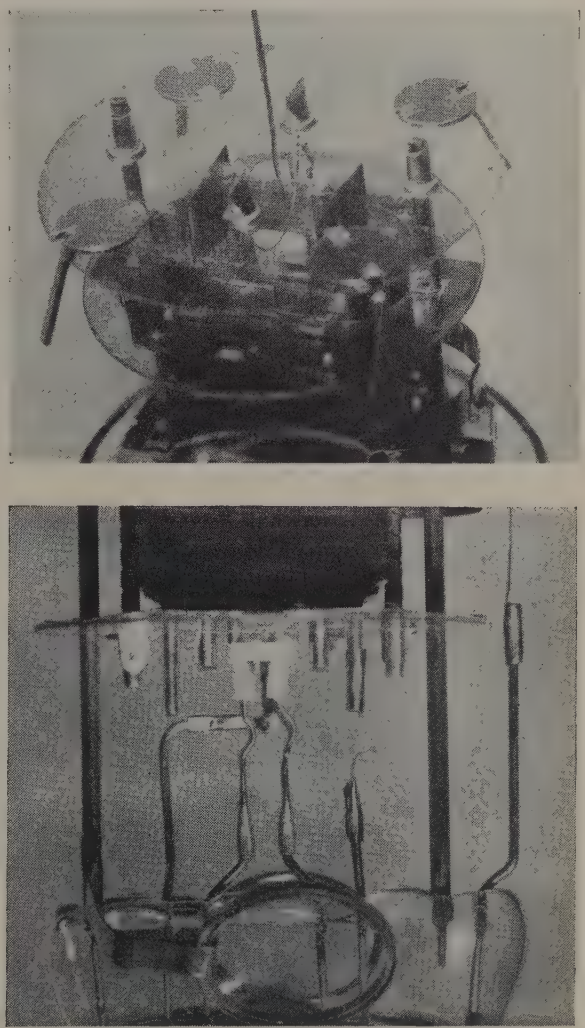


Fig. 6.—Type 10P valves—constructional details.

by two pure alumina tubes slipped over the heater limbs and at the other end by a bowed tungsten wire. An improved locking arrangement for the cathode sleeve, also shown in Fig. 6, avoids the weaknesses of conventional cathode crimps and precludes any movement between the cathode sleeve and mica insulator at the locked end.

In order to achieve a reliable structure, the number of joints has been minimized by making all connectors as integral as

possible. In general, joints have been designed as mechanical joints with a weld or welds acting as reinforcement. Nickel-silver bimetal sheet has been used for fabricating a number of the connectors, so that when welded the silver flows over the mating areas to yield effectively a brazed joint.

Despite capacitive disadvantages, the lamp-style glass pinch remains unsurpassed for reliability and is therefore used for the type 10P valves, as shown in Fig. 5. Additional support is given to the structure by means of stainless-steel 'snubbers' bearing on the glass bulb, which is made of lead glass in order to mitigate the electrolytic effect. Although the valve is fitted with a normal octal base, the pinch wires are extended through the base pins to permit direct soldering into the circuit.

Manufacture of the valve is carried out under the semi-clinical conditions now commonly used for the production of submerged repeater components. Careful visual examinations are made at various stages of manufacture followed by an intensive microscopic examination for every selected repeater valve.

(6) CONCLUSION

Whilst the type 10P valves represent an advance in the development of submerged repeater valves, it would be complacent to imagine that any degree of finality had been reached. The first valves made with tungsten-nickel cathodes have been on life test for nearly five years, and the earliest type 10P valves are well advanced into their second year. Because there is no reliable method for accelerating the life of a thermionic valve and correlating the results with normal rated performance, a complete assessment of these first deep-water repeater valves must await the passage of time.

Much of the progress reported in the paper has in fact originated from a careful examination of life trends in the context of the mechanical and production history of the valves. Clear lines of action, which have led to improved performance, have often been indicated after comparatively short testing periods. It may well be that continued vigilance will point the way to still further refinements of both design and manufacturing technique.

(7) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in the paper.

The authors would also like to acknowledge the assistance given by their colleagues in the Thermionics Group of the Post Office Engineering Department and, in particular, the helpful advice and criticism they received from Dr. G. H. Metson, who has guided the work throughout. In addition the authors would like to thank Mr. E. G. Rowe, of Standard Telephones and Cables, Ltd., for help given during the later stages of the development of the valves.

(8) REFERENCES

(1) McNally, J. O., Metson, G. H., Veazie, E. A., and Holmes, M. F.: 'Electron Tubes for the Transatlantic Cable System', *Proceedings I.E.E.*, Paper No. 2279 R, January, 1957 (**104** B, Suppl. 4, p. 60).

(2) Brockbank, R. A., Walker, D. C., and Welsby, V. G.: 'Repeater Design for the Newfoundland–Nova Scotia Link', *ibid.*, Paper No. 2249 R, January, 1957 (**104** B, Suppl. 4, p. 91).

(3) Metson, G. H., and Holmes, M. F.: 'Deterioration of Valve Performance due to Growth of Interface Resistance', *Post Office Electrical Engineers' Journal*, 1953–54, **46**, p. 193.

(4) Eisenstein, A., John, H., and Affleck, J. H.: 'Double Activators in Oxide-Cathode Base Alloys', *Journal of Applied Physics*, 1953, **24**, p. 631.

(5) Holmes, M. F.: 'The Relationship between Cathode Emission, Cathode Resistance and Mutual Conductance in Receiving Valves', *Proceedings I.E.E.*, Monograph No. 214 R, December, 1956 (**104** C, p. 251).

(6) Metson, G. H.: 'The Conductivity of Oxide Cathodes. Part 5.—Functional Structure of the Cathode', *ibid.*, Monograph No. 289 R, February, 1958 (**105** C, p. 374).

(7) Baker, B. O.: 'Gold as a Grid Emission Inhibitor in the Presence of an Oxide-Coated Cathode', *British Journal of Applied Physics*, 1953, **4**, p. 311.

(8) Metson, G. H., Wagener, S., Holmes, M. F., and Child, M. R.: 'The Life of Oxide Cathodes in Modern Receiving Valves', *Proceedings I.E.E.*, Paper No. 1137 R, 1951 (**99**, Part III, p. 69).

(9) Reynolds, F. H., Johnson, C. B., and Rogers, M. W.: 'Growth of Anode-to-Grid Capacitance in Low-Voltage Receiving Valves', *ibid.*, Paper No. 2384 R, September, 1957 (**104** B, p. 487).

(10) Child, M. R.: 'Notes on a Source of Intermittent Noise in Oxide-Cathode Receiving Valves', *ibid.*, Paper No. 2140 R, September, 1956 (**103** B, p. 667).

(11) Becker, J. A.: 'Phenomena in Oxide-Coated Filaments', *Physical Review*, 1929, **34**, p. 1323.

(12) Metson, G. H.: 'A Study of the Long-Term Emission Behaviour of an Oxide-Cathode Valve', *Proceedings I.E.E.*, Paper No. 1790 R, April, 1955 (**102** B, p. 657).

(13) Metson, G. H., Rickard, E. F., and Hewlett, F. M.: 'Some Experiments on the Breakdown of Heater-Cathode Insulation in Oxide-Cathode Receiving Valves', *ibid.*, Paper No. 1857 R, September, 1955 (**102** B, p. 678).

(14) Reynolds, F. H.: 'Design of Pentode Receiving-Type Valves', Electronic Engineer's Reference Book (Heywood, 1958), Sections 424–427.

(15) Liebmann, G.: 'The Calculation of Amplifier Valve Characteristics', *Journal I.E.E.*, 1946, **93**, Part III, p. 138.

(9) APPENDICES

(9.1) Specification for L2 Alloy

The alloy shall be prepared by powder-metallurgical technique from carbonyl-process nickel and tungsten powder. The tungsten content shall be $3{\cdot}5\% \pm 0{\cdot}2\%$ by weight. The weight of nickel and tungsten together shall be not less than 99·90%. The percentage of individual impurity elements shall not exceed the following values:

Iron	0·03%
Carbon	0·02%
Sulphur	0·003%
Silicon, Aluminium, Titanium, Magnesium, Chromium, Manganese and Copper	each 0·005%

(9.2) Abridged Specifications

(*a*) *Type 10P1 Valve.*

Under conditions $V_a = V_{g2} = 70$ volts, $V_h = 10{\cdot}2$ volts $I_a = 10$ mA.

	Minimum	Maximum
Mutual conductance, mA/V	6·3	7·4
Working-point, volts	−2·8	−4·2
Screen-grid current, mA	1·8	2·7
Anode impedance, kilohms	150	
Input capacitance, pF	10·8	12·7
Margin between working-point and cut-off ($I_a = 0{\cdot}1$ mA), volts		5·0

	Minimum	Maximum
Margin between working-point and onset of positive grid current (I_{g1} = 0·1 μA), volts	2·0	
Reverse grid current, μA		0·1
Heater current, mA	282	319
With valve cold.		
Anode-grid capacitance, pF	0·020	0·032
Output capacitance, pF	6·2	7·2

(*b*) *Type* 10*P*2 *Valve.*

Under conditions V_a = 40 volts, V_{g2} = 50 volts, V_h = 10·2 volts, I_a = 3 mA.

	Minimum	Maximum
Mutual conductance, mA/V	5·4	6·7
Working-point, volts	−1·1	−2·1
Screen-grid current, mA	0·55	0·85
Anode impedance, kilohms	250	
Input capacitance, pF	12·9	15·1
Margin between working-point and cut-off (I_a = 0·1 mA), volts		2·0
Margin between working-point and onset of positive grid current (I_{g1} = 0·1 μA), volts	0·4	
Reverse grid current, μA		0·1
Heater current, mA	282	319
Equivalent noise resistance, ohms ..		1 000
With valve cold.		
Anode-grid capacitance, pF	0·015	0·025
Output capacitance, pF	6·8	7·6

621.385.3:621.396.97

The Institution of Electrical Engineers
Paper No. 3200 E
Mar. 1960
©

HIGH-POWER TRANSMITTING VALVES WITH THORIATED FILAMENTS FOR USE IN BROADCASTING

By H. S. WALKER, M.B.E., W. H. ALDOUS, B.Sc., Associate Members, R. G. ROACH, B.Sc., J. B. WEBB, D.F.H., Associate Member, and F. D. GOODCHILD, B.Sc.(Eng.), Member.

(*The paper was first received* 22*nd June, and in revised form* 7*th November,* 1959.)

SUMMARY

A major change that has occurred in high-power transmitting valves during the last decade is the introduction of the thoriated tungsten filament or cathode. The paper describes the technology of these cathodes and some of the valve characteristics resulting from the change-over from a pure tungsten cathode. Details are given of the various properties of thoriated tungsten cathodes resulting from the carburization process which is necessary to give stable emission. The performance is given of six typical types of transmitting valves used both in the sound and television services of the B.B.C. These are typical of many other types of large valves with thoriated cathodes which differ only in minor aspects from those described. Detailed performance results are given, together with a study of the economies of thoriated cathodes in practice. In considering the installation of these valves in place of the valves using bright tungsten cathodes, consideration has been given to the economic aspects of installation. Illustrations of the six types of valves described in the paper and survivor curves both for tungsten and thoriated tungsten valves are given, together with the curves showing loss of programme due to valve failure.

It seems clear from the results that not only do thoriated tungsten filaments or cathodes give economy in operation, but the lives in many cases are considerably extended.

(1) HISTORICAL NOTES

The discovery of the enhanced emission obtainable from tungsten filaments containing a trace of thoria was first made by Langmuir as early as 1914, but the first description of the phenomenon was not given by him until about 1923. Work in this connection had been carried out by several others towards the end of the First World War, and by 1922–23 several types of valve using a thoriated tungsten filament made their appearance. However, satisfactory valves were not produced until the process of carburizing the outer layer of the filament had been developed by heating it in a hydrocarbon vapour at about 2300° K. This converts the outer skin of the filament to tungsten carbide, which is able to reduce the thoria in the operating range of temperature (1900–2100° K). The emission from a carburized filament is much more stable and less subject to vacuum conditions than an uncarburized filament, and is dealt with in more detail elsewhere in this paper. Some of the first small receiving valves, such as the DEQ and DEV, were produced with carburized thoriated filaments. These were followed fairly quickly by the LS5, which was a larger receiving valve and could, under suitable circumstances, be used as a transmitting valve dissipating up to about 10 watts at its anode at about 400 volts. This valve remained very popular for a number of years. Efforts were made to develop larger valves, and by the early 1930's valves capable of dissipating 250 watts at their anodes and by the late 1930's valves dissipating up to 1·25 kW with voltages of up to 3 kV became a practical proposition. It was soon apparent, however, that these valves had to be made particularly free from gas, because any ions present could bombard the filament and soon destroy its emission. The higher the anode voltage, of course, the more damaging did this effect become. At the commencement of the Second World War high-power valves were required for pulsed radar transmitters, and the cathodes had, of necessity, to have a large emission. Since the emission was required only for a fraction of the operating cycle, and for the majority of the time reactivation could take place, thoriated tungsten filaments proved very satisfactory, even with anode voltages of up to 40 kV. Valves for pulse operation were made both with silica and metal/glass envelopes and gave good lives. Large valves for continuous operation at high voltages could not be made until the development of continuously-acting getters, one of the most useful of which is zirconium. Since about 1948, large valves of 15 kW anode dissipation upwards have been made, operating with carburized thoriated tungsten filaments, and it is the purpose of the paper to survey the present situation and to give the results of the operation of these valves in practice, together with a study of the savings in operating power, particularly so far as filaments are concerned, compared with the more conventional type of large valve fitted with a pure tungsten filament.

The period since 1945 has also seen extensive development in radio transmission in the very-high-frequency field. Valves for this service require to be as small as possible in order to reduce inter-electrode capacitances and lead inductances, for which the thoriated tungsten filament is very advantageous. In fact, many of the most modern types of valves could not have been designed using a bright tungsten emitter. The development of these new types with thoriated tungsten filaments and the use of special grid material has led to many problems of mechanical design, some of which have become apparent only during the actual operation of valves in the field. There is no doubt that all these problems can be overcome in the course of time, and that all the new designs will be free from mechanical troubles and will run to the full emission life of the filament.

(2) TECHNOLOGY OF THORIATED TUNGSTEN FILAMENTS

(2.1) Reasons for Carburization

The addition of about 1% of thoria (ThO_2) to tungsten was originally made in order to modify the recrystallization properties of the metal. Langmuir[1] found that, by suitable heat treatment in vacuum, some of this thoria could be reduced to thorium which diffused to the surface, and the resultant surface monatomic film gave much enhanced thermionic-emission properties to the wire. This discovery led to the adoption of thoriated tungsten as a filamentary cathode, and a number of valve types

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
This is an 'integratmg' paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.
Mr. Walker and Mr. Webb are with the British Broadcasting Corporation.
Mr. Aldous is at the Research Laboratories of The General Electric Company, Limited.
Mr. Roach is with the English Electric Valve Co., Ltd.
Mr. Goodchild is with Standard Telephones and Cables, Ltd.

were, in fact, successfully manufactured using the Langmuir process, both in this country and in America. At that time, however, it was not easy, with the techniques available, to obtain the degree of vacuum in the valves required to maintain stable emission.

Thoriated tungsten was really established as a commercially useful electron emitter only by the introduction of the carburizing process. This also appears to have arisen accidentally, in that carburization of lamp filaments, which occurred during processing in the residual pressure of oil vapour from the mechanical vacuum pumps, was being investigated at the same period. The fact that more stable emission was obtainable from carburized thoriated tungsten was observed by Langmuir and mentioned by him at the end of his first paper on the subject.

Concurrent work by Andrews[2] on the effects on emission of deposited films of thorium was misinterpreted to prove that the rate of evaporation of thorium from a carburized filament was lower than that from an uncarburized filament. This incorrect conclusion has been repeatedly quoted in subsequent literature as being a major factor in the improved emission stability given by carburization. In fact, a recent direct comparison[3] has shown that, at a given temperature, when the filaments are fully activated, the rates of evaporation of thorium from carburized and uncarburized filaments are equal, to the first order. The rates of supply of thorium to the surface layer from within the filament must therefore also be approximately equal, since there is no accumulation of thorium on the surface. A possible reason for the improved stability of emission of the carburized filament under imperfect vacuum conditions could be the break-up of the surface by the carburization. This gives more grain boundaries along which diffusion can take place, and therefore shorter migration distances across the surface to maintain uniform coverage.

During its life a carburized filament has the additional advantage that there is a continuous reaction between carbon and thoria at the operating temperature of 2000° K, thereby maintaining a sensibly constant supply of thorium in the filament. The uncarburized filament has to be flashed at 2800° K to promote reduction of thoria to thorium by the tungsten. Subsequent reduction at 2000° K is negligible, so that the usable reservoir of thorium built up during the flashing operation will steadily diminish throughout life.

(2.2) Methods of Carburization

The carburization of tungsten filaments may be carried out in a number of ways:

(*a*) By coating the surface with the required amount of carbon, and subsequently heating the coated wire in vacuum. The main difficulty in the application of this method lies in the sudden decrease of emissivity, and therefore rise of temperature, which occurs when the carbon in any area has been completely absorbed.

(*b*) By embedding the filament in bulk carbon and heating in a furnace. Partial carburization to a uniform depth in this manner is not easy.

(*c*) By electrical heating of the filament in the presence of a hydrocarbon gas or vapour.

Method (*c*) is the one usually employed. It may be carried out on the assembled valve during the pumping operation by the temporary introduction into the vacuum system of a solid with a low vapour pressure, such as naphthalene, or of a low pressure of a gas such as acetylene. It is more common, however, to precarburize, either as single filaments or as complete filament assemblies, in hydrocarbon vapour carried by hydrogen which has passed through a carburettor containing the liquid hydrocarbon, e.g. benzene, xylene, etc.[4] The filament temperature during this carburization is about 2300° K, and, owing to the cooling effect of the hydrogen, the heating power required is approximately four times the final running power *in vacuo*. The large powers involved make the use of water-cooled enclosures necessary. Carburization is not a critical procedure, the range of time and temperature to obtain a given result being fairly wide. The controlling factors are the rate of decomposition of hydrocarbon at the filament surface and the rate of diffusion of carbon into the tungsten. The rate of decomposition of hydrocarbon at the filament surface does not depend to any large extent on the temperature. Indeed, there is evidence that, when using naphthalene, every molecule reaching the hot surface is decomposed, so that the rate of decomposition depends only on pressure. When the hydrocarbon is in excess of hydrogen, this may not hold in detail, but filament temperature is still not a major factor. The rate of diffusion of carbon into the tungsten increases rapidly with rise of temperature. For any given partial pressure of hydrocarbon there is therefore a minimum temperature to use, below which carbon will build up on the filament surface, with consequent additional cooling by radiation, further drop in temperature and more deposition.

To carburize successfully, it is essential that water vapour and oxygen should be kept away from the hot filament. Thus condensation on the walls of the water-cooled enclosure must be avoided, constructional materials which give off water vapour when in contact with hot gases must not be used, and an adequate seal against the ingress of air must be provided. The degree of carburization used may depend on the wire size and on the valve in which it is used. As discussed in Section 2.4 the emission life is determined by the loss of carbon, so that a given fractional depth of carburization will give a longer life with a large-diameter wire than with a small one. Figures used commercially usually vary between 10 and 30% of the cross-sectional area, although Ayer[5] shows cross-sections of filaments with up to 40% of the area carburized. The process is usually carried out with the filament electrically heated at either constant voltage or constant current, and, in either case, the depth of carburization is controlled by the increase in filament resistance that occurs. The accuracy of this control will depend on the maintenance of uniform surface finish—and therefore uniform radiating properties—on the wire used. In one method of manufacture, a greater cross-sectional area is carburized than is finally required, and the heating is then continued for a short time in pure hydrogen to decarburize a shallow surface layer. A photomicrograph of the cross-section of such a filament is shown in Fig. 1.

(2.3) Effects of Carburization on the Properties of Tungsten Filaments

(2.3.1) Composition.

The changes in chemical composition of tungsten filaments which occur during gaseous carburization have been described by Horsting.[4] The phase diagram deduced from his and earlier work, which has been slightly modified by Norton,[6] shows that the part of the metal which is usually referred to as uncarburized is, in fact, a solution of about 0·05% of carbon in tungsten. The carburized layer formed under the correct conditions, i.e. without surface deposition of carbon, consists of di-tungsten carbide (W_2C) which has some tungsten dissolved in it. The existence of this definite phase gives rise to the sharp demarcation line as the diffusion goes inward. The solubility of tungsten in di-tungsten carbide becomes less as the temperature is lowered, and some therefore precipitates on cooling. This gives rise to a characteristic laminar structure which often shows in cross-sections of cold filaments (Fig. 1).

The specific gravity of di-tungsten carbide is less than that of tungsten, so that the process of carburization will increase the volume of a filament, and therefore its diameter. The specific

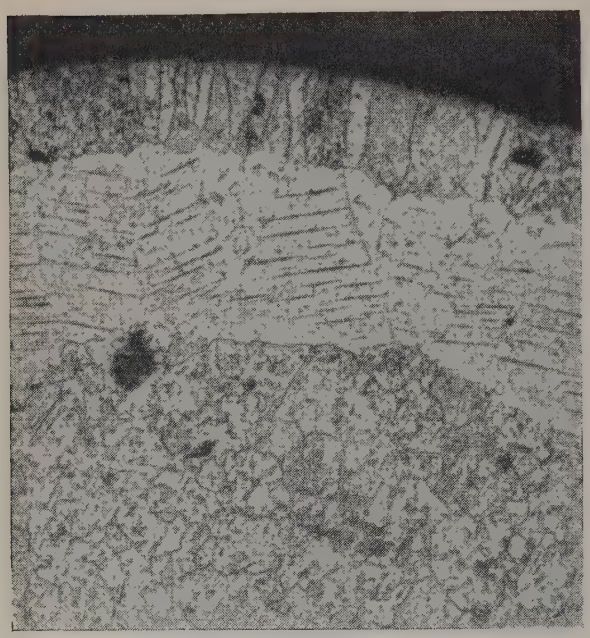


Fig. 1.—Photomicrograph of filament section showing decarburized surface.

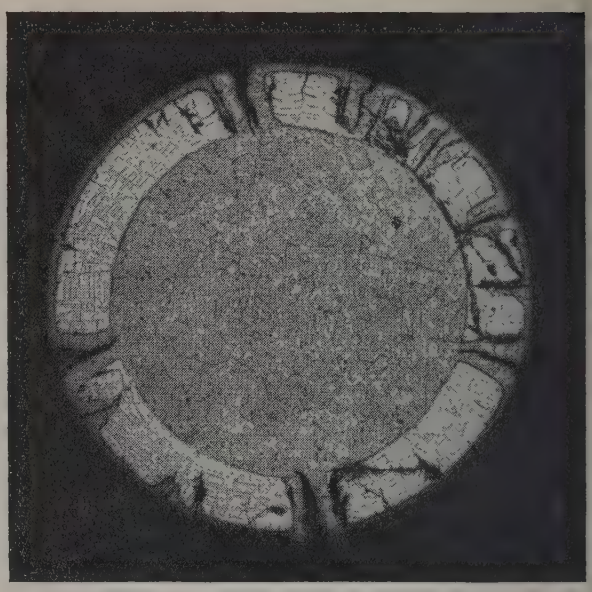


Fig. 2.—Photomicrograph of split wire.

gravity of the metals,[7] especially when thoriated, depends to some extent on the details of manufacture and processing, but taking figures of 19·0 for thoriated tungsten and 17·0 for thoriated tungsten carbide, a filament after carburization of 25% of its cross-sectional area will increase in volume by 3%, or, if this increase is all assumed to take place radially, will increase in diameter by 1·5%. This increase can, under some conditions, form a useful indication as to whether wires have been carburized.

(2.3.2) **Thermal Properties.**

The thermal expansions of di-tungsten carbide and tungsten are given in the literature as follows:

Tungsten carbide[8] (20–2400° C): $1{\cdot}2 \times 10^{-6}$ per deg C, in the *a*-axis direction.
$11{\cdot}4 \times 10^{-6}$ per deg C, in the *c*-axis direction.
Tungsten[9] (at 20° C): $4{\cdot}5 \times 10^{-6}$ per deg C.
(20–2400° C): $5{\cdot}8 \times 10^{-6}$ per deg C.

The large amount of anisotropy in the carbide crystals and the difference between the mean expansion of the carbide and that of tungsten will lead to the development of internal stresses in the wire on temperature cycling, and, if there is any tendency to split in the original wire, these will probably open up (Fig. 2) as a result of carburization.

The thermal emissivity of filaments with a tungsten-carbide surface is 20% higher than that for uncarburized tungsten filaments.

The melting point of the eutectic between tungsten and di-tungsten carbide is 2750° K, compared with 3680° K for pure tungsten and 3130° K for di-tungsten carbide. This lower melting point is, however, still far above the usual operating temperature range of thoriated tungsten filaments (1950–2050° K).

(2.3.3) **Mechanical Strength.**

Although no precise data on the bending strength of carburized wires are available, general experience indicates that they are more brittle than similar uncarburized wires. They are, however, adequately strong for use in large transmitting valves, where protective measures for transit must be taken in any case. The brittleness decreases rapidly with rise of temperature and is negligible at a few hundred degrees Celsius.

(2.3.4) **Electrical Resistance.**

The resistivity[6] of di-tungsten carbide is 80×10^{-6} ohm-cm at 293° K and 119×10^{-6} ohm-cm at 2000° K, while that of tungsten[10] is $5{\cdot}5 \times 10^{-6}$ ohm-cm at 293° K and $55{\cdot}7 \times 10^{-6}$ ohm-cm at 2000° K. From these figures the increase in resistance of a wire as a function of the percentage of the original cross-sectional area which has been carburized may be determined. Thus if 25% of the original area is carburized, the hot resistance will increase by 15%, and the cold resistance by 30%.

(2.3.5) **Electron Emission.**

Carburization has no measurable effect on the magnitude of the electron emission obtainable from thoriated tungsten at a given temperature.[7] It does, however, markedly affect the ease with which the emission is obtained. The high-temperature flash at 2800° K described by Langmuir is no longer necessary; it would, in fact, be harmful, since it is above the melting point of the tungsten–tungsten-carbide eutectic. Normally the slight over-voltage the filament receives during the pumping operation is quite adequate for the carbon to reduce sufficient thoria to obtain the required monolayer of thorium on the wire surface, and therefore the emission. The saturated emission obtainable is of the order of 3–4 amp/cm^2 at 2000° K, with an emission efficiency of 100–120 mA/watt.

(2.4) Effects during Life

(2.4.1) **Loss of Emission.**

The emission from thoriated tungsten depends on the maintenance of a monatomic film of thorium on the surface. This thorium arises from the reduction of thoria by carbon, and the loss of emission which occurs at the end of life could therefore be due to loss of either thoria or carbon. All experience with glass-envelope valves indicates that it is the carbon which is lost first, adequate thoria still being present in the filament. Experience with large cooled-anode valves shows that their emission lives are longer than those obtained from glass-envelope valves,

but that, unless the life is about 100 000 hours or more, it will probably still be terminated by the loss of carbon.

(2.4.2) **Loss of Carbon.**

With tungsten containing 1% of thoria, the amount of carbon necessary to give complete reduction of the thoria would be obtained by converting a little above 3% of the cross-sectional area to di-tungsten carbide. As it is usual to carburize between 10 and 30% of the cross-sectional area, the majority of the carbon must be lost at the surface by evaporation or chemical reaction with residual gases. Figures for this rate of loss have been given by Ayer,[5] and are plotted in Fig. 3 for comparison

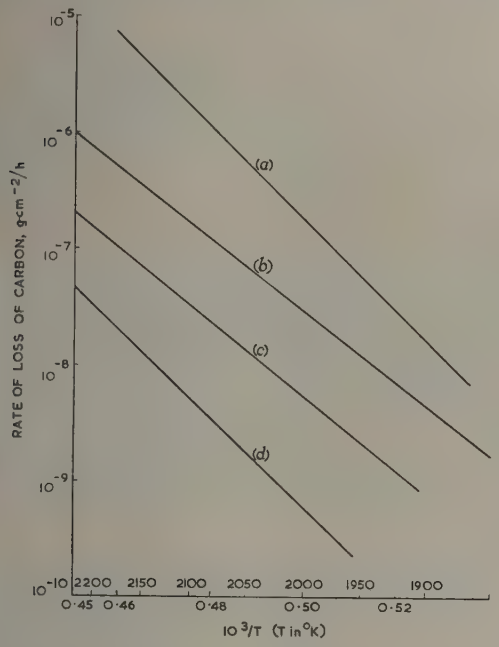


Fig. 3.—Loss of carbon as function of temperature.

(*a*) Ayer. (*c*) Wright (10^{-6} mm H_2O).
(*b*) Wright (10^{-5} mm H_2O). (*d*) Carbon.

with data obtained by Wright[11] under controlled conditions, and with the rate of evaporation of carbon from bulk carbon.

Ayer's figures, which were obtained from operational experience with glass-envelope valves, almost certainly underestimate the life that can be obtained from modern cooled-anode valves, where the vacuum conditions are better. The figures indicate that a 1 mm filament, with 25% of its original cross-section carburized, would lose all its carbon in about 35 000 hours when run at 2 000° K.

Wright's figures refer to conditions without anode current flow, and therefore without positive-ion bombardment of the emitting surface. They may therefore underestimate the rate of decarburization for given vacuum conditions. It is known in practice that a shorter emission life is obtained when a given type of valve is run at a higher mean cathode current. True figures for the mean life, as determined by carbon loss, must therefore await further operational experience.

The location as well as the amount of tungsten carbide in the filament will change during life, since both are dependent on the rate of loss of carbon and on the rate of its diffusion through the metal. Previous published work has usually dealt with cases where the emission life was relatively short. In these cases the rate of carbon loss had the greatest effect, and, in much the same way as in surface decarburization during processing (Fig. 1), the carbon disappeared mainly from the outer surface of the original carbide layer.

For the long-life high-power valves, where the rate of loss of carbon is low, the picture is quite different. The diffusion process is now predominant, and it is found that a distribution of carbide particles is produced which is roughly uniform across the section of the wire. The photomicrograph of Fig. 4 shows

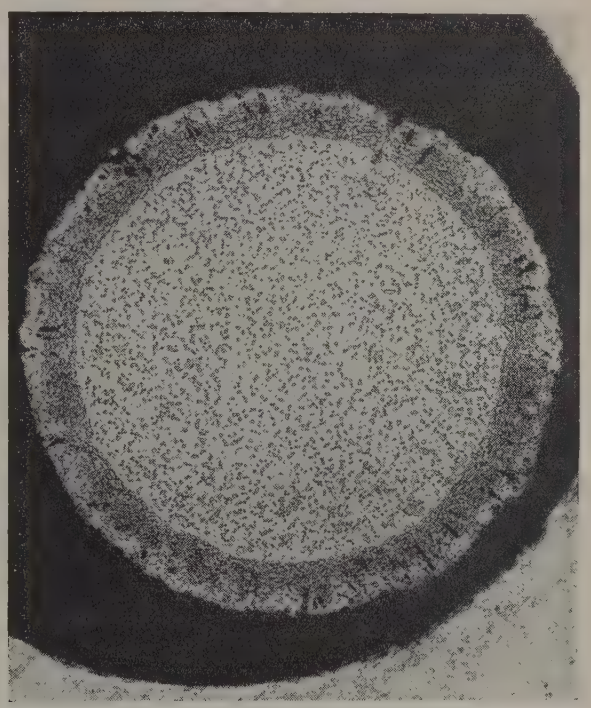


Fig. 4.—Photomicrograph of filament showing carbide disseminated as particles over cross-section of wire.

this for a filament from a valve which still had good emission after running for 7 000 hours. In this case the inward diffusion has not gone to completion, the concentration of carbide in the originally carburized layer still being considerable.

The change in electrical resistance of the filament during life will be affected by the pattern of the tungsten-carbide location. If the carbide has a tubular section, occupying a fraction, p, of the total volume, then, neglecting the effects of the change in volume on carburization, the effective conductivity, σ_1, of the composite metal will be given in terms of σ_w, the conductivity of tungsten, and σ_c, the conductivity of tungsten carbide, by

$$\sigma_1 = \sigma_w - p(\sigma_w - \sigma_c) \quad . \quad . \quad . \quad . \quad (1)$$

Where $p = 0{\cdot}25$ initially and the carbide section remains tubular through life with loss of carbon at a uniform rate, the change in effective room-temperature conductivity with time will be linear as shown by Fig. 5 [curve (*a*)].

However, if the carbide becomes uniformly disseminated as small particles through the tungsten, the effective conductivity, σ_2, will be given by[12]

$$\sigma_2 = \frac{2\sigma_w + \sigma_c - 2p(\sigma_w - \sigma_c)}{2\sigma_w + \sigma_c + p(\sigma_w - \sigma_c)}\sigma_w \quad . \quad . \quad . \quad (2)$$

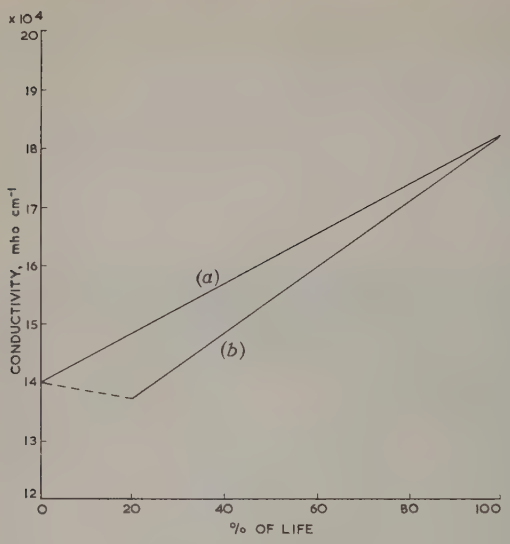


Fig. 5.—Change of effective conductivity during life.
(*a*) Tubular carbide section. (*b*) Disseminated carbide.

For $p = 0{\cdot}25$ initially, and the change to uniform dissemination taking place in the first 20% of the life, the effective conductivity during life is given by Fig. 5, curve (*b*).

These curves show that, without knowledge of the disposition of the carbide in the filament, it is not possible to use resistance measurements made during early life to determine the rate of loss of carbon, and therefore to predict the probable ultimate emission life of a valve.

(2.4.3) **Loss of Thoria.**

The work of Jenkins and Trodden[3] shows that, at a running temperature of 2000° K, the rate of loss of thorium from a carburized filament is about 7×10^{-12} g-cm^{-2}/sec. If this is assumed to remain constant throughout life, then, for a 1 mm filament, the time for complete loss of the original 1% of thoria will be about 150 000 hours.

(2.5) Effects of Thoriated Tungsten Cathodes on the Properties of Valves

(2.5.1) **Filament Characteristics.**

The rate of change at operating temperature of many parameters of pure tungsten filaments for small changes of voltage or current is well known (see, for example, Reference 13). The corresponding approximate figures for the free-space characteristics of carburized thoriated tungsten at 2000° K are as follows: A 3% increase in applied voltage produces 2% increase in current, 5% increase in power, 1% increase in temperature, 20% increase in emission, and 50% decrease in life due to carbon loss. In practice, some of these figures may be modified a little by heat reflection from surrounding electrodes and by the cooling effects of filament supports.

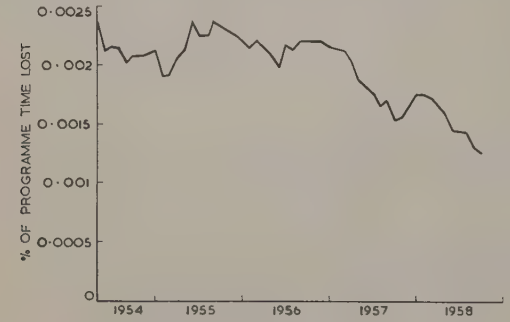


Fig. 6.—Loss of programme due to valve failure.

(2.5.2) **Grid Emission.**

When using pure tungsten filaments, the major limitation to the power output from a given size of valve is the cathode emission, and the rest of the structure is designed to suit this. With thoriated tungsten filaments, however, the distillation of thorium on to the grid surfaces reduces their work function so much that primary grid emission often becomes the limiting factor in design.

In valves with pure tungsten filaments, the internal metal surfaces may be made very clean and reproducible during the processing, and primary grid emission can occur only at very high temperatures. For example, for tungsten and molybdenum, the most commonly used grid materials, the temperatures required to give an emission of 1 μA/cm^2 are about 1 600° K and 1 500° K, respectively. When these metal surfaces are covered with thorium, it is possible to obtain the same emission at about 1 100° K.

In practice, some grid emission can be tolerated in a large transmitting valve. However, owing to the combined effects of radiation from the filament and direct electron bombardment, the grid may have to operate at a temperature where emission from tungsten or molybdenum would become excessive, and some suitable treatment of the grid surface must be used to reduce the emission. A common treatment is to use a sheath of platinum applied during wire manufacture, while coatings of zirconium or titanium or their compounds are also used. In these ways the temperature for a given grid emission can be raised by at least 300° K.[13]

(2.5.3) **Anode and Grid Currents.**

The thoriated-tungsten transmitting valve is always operated with a peak cathode current several times less than the saturated cathode emission. The result is that the effect of saturation on the valve characteristics is negligible, and in this respect is different from the tungsten-filament valve. The latter valve commonly shows saturation from some parts of the filament at a total current well below the overall saturation level. As a result, characteristics of tungsten-filament valves cannot be deduced from a few parameters, such as amplification factor, mutual conductance, etc. On the other hand, the characteristics of valves with thoriated tungsten filaments follow simple theoretical considerations much more closely, and the characteristics can be deduced from basic parameters to a useful extent. Provided that the space-charge law is obeyed, which it will be so long as no part of the cathode is saturated, the cathode current of a triode valve is related to the voltages on the anode and grid by the formula

$$I_k = \mathscr{P}(V_g + V_a/\mu)^{3/2} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $\mathscr{P}$, the perveance, and $\mu = \partial V_a/\partial V_g$, the amplification factor, are basic constants of the triode and are independent of the voltage and current. The constant μ is familiar to all valve users. For triodes its value lies in the range 5–100. The constant $\mathscr{P}$, although basic and often quoted for the electron guns used in beam devices such as klystrons, may not be familiar to some users. It is usually given in micropervs (1 perv = 1 amp/volt$^{3/2}$), and a typical figure for a 20 kW triode would be in the region of 1 000 μp.

It has been customary to specify triodes by the values of the

amplification factor, μ, and the mutual conductance, $g_m = \partial I_a/\partial V_g$), although the latter is far from constant and may, in some parts of the operating range, differ by a factor of at least two from the manufacturer's quoted value. In the negative grid region none of the cathode current will go to the grid, and g_m may be obtained by differentiating eqn. (3) to give

$$g_m = \tfrac{3}{2}\mathscr{P}^{2/3} I_a^{1/3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

which shows how g_m varies with anode current when $\mathscr{P}$ remains constant.

In the positive grid region some of the cathode current is intercepted by the grid and is lost to the anode. The division of current between anode and grid may be defined by a third constant, the current division ratio, δ, given by

$$\delta = \left(\frac{I_a}{I_g}\right) V_a = V_g \quad . \quad . \quad . \quad . \quad . \quad (5)$$

which is not dependent upon the absolute values of V_a and V_g. When V_a is greater than V_g the division ratio over the range $V_g < V_a < 4V_g$ may be estimated from

$$\frac{I_a}{I_g} = \delta\left(\frac{V_a}{V_g}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which applies over those regions of the characteristics where the major part of the grid current arises.

It will usually be found that direct substitution of thoriated tungsten for pure tungsten in a given valve design will, if anything, increase the ratio of grid current to anode current for given grid and anode voltages. However, the lack of saturation will normally mean that the required anode current can be obtained with less positive grid drive, so that less drive power will be required for a given maximum output power.

One other more subtle difference is shown in the grid-current characteristics. The grid current for small positive grid voltages rarely becomes negative, as it often does with pure tungsten filaments. This is because the thoriated tungsten filament has more emission from its cold ends, and the space charge from this prevents the escape of secondary electrons from the grid to the anode.

(2.5.4) **Switching Transients.**

The cold resistance of a carburized thoriated tungsten filament is about one-eighth of its hot resistance. When switched in the cold state on to a low-impedance supply, heavy currents will flow and the magnetic forces between filament limbs may cause distortion. This is a known problem with pure tungsten filaments, and is dealt with by multi-step switching using series resistances, or other suitable means of current limiting. For a given diameter of filament, the combined effect of lower running current and of the lower ratio of hot to cold resistance gives a surge current with carburized thoriated tungsten which is only about 40% of that obtained with pure tungsten, and is therefore dealt with more easily. With a.c. supplies a current-limiting transformer may be used. If a switched series resistance is used, this may conveniently be a single resistance whose value is 70% of the hot filament resistance and which is short-circuited after 20 sec.

(2.5.5) **Hum.**

When directly heated filaments are operated on alternating current, the anode current is modulated both by the electric and magnetic fields of the filament. The magnetic component is proportional to the square of the filament current per limb.

For a given geometrical design of filament and a given working anode current, the increased emission per unit area obtained by changing from tungsten to thoriated tungsten would enable the filament diameter and therefore the current per limb to be reduced. For other reasons, usually that of mechanical strength, this diameter reduction is not always incorporated. However, the lower operating temperature will still give lower current for the thoriated tungsten filament, and, in practice, the hum level is usually reduced by about 4–5 dB by the change from pure tungsten.

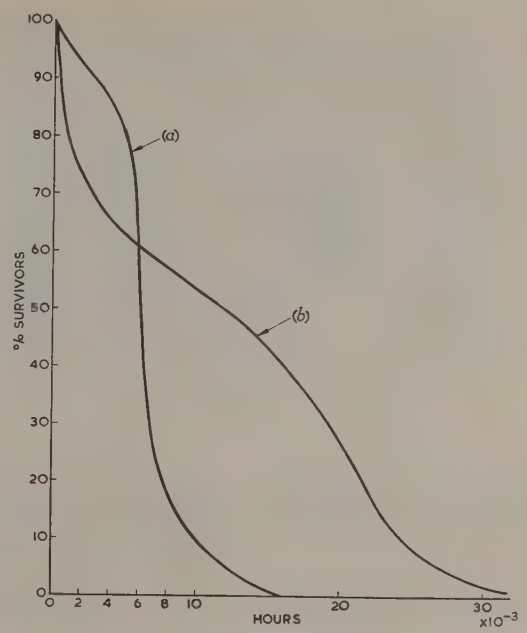


Fig. 7.—Survivor curves for (*a*) bright tungsten and (*b*) thoriated tungsten filament valves.

(2.5.6) **Circuit Currents.**

Theoretical considerations indicate that the r.f. currents flowing through the thoriated filament in a valve could have a marked effect on its life if the valve is loaded to the full extent allowed by the emission. As an example, consider a particular tetrode with a grid-cathode capacitance of 50 pF, of which approximately half is active capacitance, operating at a frequency of 200 Mc/s, with a 400-volt r.m.s. swing applied to the input. Under these conditions the r.m.s. capacitive current flowing between grid and cathode will be 25 amp and the r.m.s. space current will be about 6 amp. Because of skin effect, the combined action of these r.f. currents, acting in addition to the 175 amp heating current, will be to raise the temperature of the filament system at one end by about 50° C. Since the rate of evaporation of carbon changes by about 2% per deg C at the operating temperature, the presence of the r.f. current should appreciably affect the life. There is some evidence that this can be detected in valves used in the broadcasting service at frequencies greater than 100 Mc/s.

(3) HIGH-POWER TRANSMITTING VALVES IN USE IN THE B.B.C. SERVICE

High-power transmitting valves were last fully surveyed before The Institution in 1938 by Bell, Davies and Gossling.[15] While the general principles of construction have remained the same, many modifications and improvements have been made, particularly in improving the robustness of the valves, and in designing them for

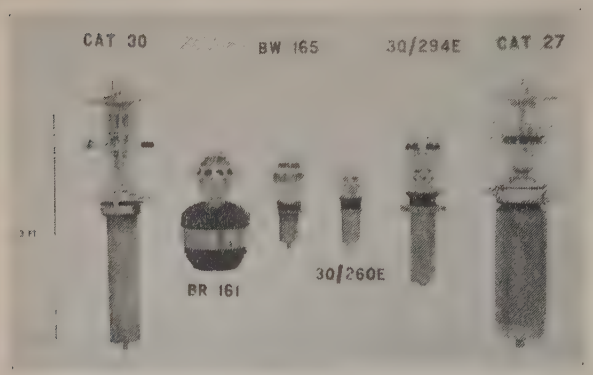


Fig. 8.—Illustrations of six types of thoriated-filament transmitting valves.

use at much higher frequencies. Fig. 8 shows six types of large transmitting valves using thoriated tungsten filaments which are in use in the B.B.C. service. A number of other types of large thoriated-filament valves are also in use, and although the following Sections are confined to the six types of Fig. 8, for which the maximum information is available, the authors have every reason to believe that the others will give comparable results in practice.

(3.1) General Description of Six Types of Large Transmitting Valves in Use at Present in the B.B.C. Service

(3.1.1) CAT27 and 3Q/294E.

The CAT27 and 3Q/294E are considered together as they have both been used by the B.B.C. as substitutes for the bright-tungsten-filament CAT17C valve which has been used for many years in double-ended transmitters operating class-C anode-modulated radio frequency up to 26 Mc/s. They are very different in size and a comparison of their performance is of considerable interest. The CAT17C has a filament power of 14·7 kW (32 volts at 460 amp) and the two substitute valves are shown in the following Table, which also gives the physical dimensions:

	P_f kW	V_f volts	I_f amp	Weight lb	Overall length in
CAT27	5·0	17·5	285	76	44½
3Q/294E	2·5	12·5	197	23	26½

The CAT27 was designed to be physically interchangeable with the CAT17C. The 3Q/294E, on the other hand, takes full advantage of the higher cathode efficiency of thoriated tungsten filaments to achieve smaller physical dimensions. This valve is in use in transmitters in many parts of the world, and its small size is of considerable importance in transportation. It will be seen that the change-over involves considerable differences in first cost, the CAT27 requiring only a change in filament voltage, whereas the 3Q/294E involves major changes, principally to the water jacket. The increased cost of change-over is just about equal to the difference in first cost of the respective valves.

It will be seen that advantage is taken of the large size of the CAT27 to use rather more filament power than is necessary and to operate the filament at a lower temperature to achieve the same peak emission. It is to be expected by this means that the life of the CAT27 will be considerably longer than that of the 3Q/294E, and the increased life and less frequent maintenance changes of the longer-life valve have to be considered against the cost of the additional filament power consumed. Owing to the unusually long lives of these valves, it will be many years before an exact comparison of costs can be made.

(3.1.2) BW165.

The BW165 is used for two distinctly different purposes:

(*a*) As the main power amplifier in the high-power television transmitters at Sutton Coldfield, Wenvoe, Kirk o'Shotts and Holme Moss (Band I).

(*b*) In the oversea broadcasting service from the Ludlow transmitter, where they are used both as power amplifiers and modulators in the short-wave band.

For these services the BW165 takes the place of the American 880 type or its counterpart, the CAT21, both of which are fitted with bright tungsten filaments. They have the following filament characteristics:

	V_f volts	I_f amp
BW165	7·2	170
880	12·6	320
CAT21	12·5	320

The change-over from the tungsten-filament valves to the thoriated-tungsten BW165 is a very simple operation, the lower filament voltage being obtained usually by altering the transformer primary connections.

(3.1.3) BR161.

The BR161 is used as a power amplifier and modulator in the Daventry Third Programme transmitter in place of the bright tungsten-filament valve BR126. The two valves are directly interchangeable, except for the filament characteristics which are as follows:

	V_f volts	I_f amp
BR126	12·5	480
BR161	9	175

The change-over presents no problems, the lower voltage being obtained by tappings on the filament transformer primary.

(3.1.4) 3Q/260E.

The 3Q/260E was the first high-power valve with a thoriated filament to be used in the B.B.C. service. During alterations some years ago to the London Home Service transmitter at Brookman's Park, opportunity was taken to introduce the valve as a cathode-follower and to note the performance of a thoriated filament under high-power conditions. Although as a cathode-follower the valve is lightly loaded, it has performed admirably.

(3.1.5) CAT30.

The CAT30 is used in place of the bright tungsten valve CAT20C, which is used as a modulator in the oversea broadcasting transmitters now using the CAT27 in place of the CAT17C. The filament voltage of the CAT30 is the same as the CAT27, i.e. 17·5 volts, so that, by using a pair of CAT27 valves in place of a pair of CAT17C, and a pair of CAT30 valves in place of a pair of CAT20C, the busbar voltage of the transmitter can be reduced from about 33 volts to about 20 volts.

This overall reduction in power consumption is very considerable and is dealt with in more detail in Section 4.

(3.2) Performance Results in Service

The standard of performance required from valves used in broadcasting is very high and can be conveniently expressed in terms of reliability as a function of the programme time lost due to valve failure. This is illustrated over the past few years by Fig. 6, which includes loss of programme on all B.B.C. Services, l.f., m.f., h.f., and v.h.f. The introduction of thoriated filament valves for test has therefore been undertaken gradually so that no serious deterioration in reliability would occur owing to

unforeseen factors. Thoriated transmitting valves have now been introduced for extended life tests in short-wave, medium-wave and television transmitters. A detailed analysis of the results, both as regards life and reliability, is given below.

CAT27.—The valve has been installed in 29 operational sockets, 25 operating under class-C anode-modulated short-wave conditions. 24 of these valves are operating at $V_a = 11$ kV, $I_a = 5$ amp. One valve operates at $V_a = 11$ kV, $I_a = 7$ amp. The remaining four valves are operating as class-B linear r.f. amplifiers at medium frequency ($V_a = 16{\cdot}5$ kV, $I_a = 6{\cdot}4$ amp). The performance obtained is as follows:

Valves in use	Peak life to date	Average running period	Number of failures
29	>32 000 h	>13 000 h	1

The one failure was due to a broken filament bracing wire at a life of 3 486 h.

CAT30.—The valve has been installed in 14 operational sockets all operating as class-B l.f. amplifiers ($V_a = 11$ kV). The performance obtained is as follows:

Valves in use	Peak life to date	Average running period	Number of failures
14	>19 000 h	>11 700 h	3

G–F	Low emission	Others
—	—	100%

3Q/294E.—The valve has been installed in 24 operational sockets all operating as class-C anode-modulated at high frequencies ($V_a = 10{\cdot}6$ kV, $I_a = 6{\cdot}5$ amp). The performance obtained is as follows:

Valves in use	Peak life to date	Average running period	Number of failures
24	>15 000 h	>7 900 h	21

G–F	Low emission	Others
76%	5%	19%

3Q/260E.—The valve has been installed in six operational sockets operating as very-low-frequency cathode-followers feeding a class-B modulator ($V_a = 12$–13 kV). The performance is as follows:

Valves in use	Peak life to date	Average running period	Number of failures
6	>20 000 h	>9 100 h	10

G–F	Low emission	Others
10%	50%	40%

BW165.—The valve has been installed in 30 operational sockets, 22 operating as class-C anode-modulated h.f. amplifiers and eight operating as linear amplifiers at television Band 1 frequencies. The performance is as follows:

Operating at $V_a = 10$ kV and $I_a = 3{\cdot}5$ amp.

	Valves in use	Peak life to date	Average running period	Number of failures
Short-wave	22	>33 000 h	>15 000 h	42

G–F	Low emission	Others
12%	57%	31%

Operating at $V_a = 8$ kV and $I_a = 2{\cdot}5$ amp:

	Valves in use	Peak life to date	Average running period	Number of failures
Television	8	>22 000 h	>6 000 h	32

G–F	Low emission	Others
6%	75%	19%

BR161.—Installed in 16 operational sockets, eight operating at low frequency class-B, and eight operating at low-frequency class-C anode-modulated ($V_a = 10$ kV, $I_a = 2{\cdot}5$ amp). The performance is as follows:

Valves in use	Peak life to date	Average running period	Number of failures
16	>20 000 h	>16 000 h	20

G–F	Low emission	Others
30%	60%	10%

In existing installations using tungsten-filament valves it has been the custom to maintain a graduated schedule of filament voltage, increasing with life according to the type of valve.

The thoriated-filament valves listed have all been operated throughout life at constant voltage, but there is some evidence to favour operation at constant power as the thoriated-tungsten-filament resistance changes with life.

Table 1

SUMMARY OF PERFORMANCE

Valve type	Cooling	Number in use at one time	Frequency band	Peak life	Average running period	Failures
				hours	hours	
CAT27	Water	29	H.F.	>32 000	>13 000	1
CAT30	Water	14	V.L.F.	>19 000	>11 700	3
3Q/294E	Water	24	H.F.	>15 000	> 7 900	21
3Q/260E	Water	6	V.L.F.	>20 000	> 9 100	10
BW165	Water	22	H.F.	>33 000	>15 000	42
		8	V.H.F.	>22 000	> 6 000	32
BR161	Air	8	L.F. / V.L.F.	>20 000	>16 000	20

(3.2.1) **Survivor Curves.**

Survivor curves of a typical tungsten-filament valve and thoriated-tungsten-filament valve are shown in Fig. 7. The precipitous initial fall in the curve for thoriated tungsten filament is explained by the incidence of mechanical faults as described in Section 1. It is reasonable to conclude that, when these early catastrophic failures are eliminated by suitable design modifications, the general shape of the two curves will be similar.

(4) A STUDY OF THE ECONOMIES OF THORIATED FILAMENTS IN PRACTICE

In existing installations using plain tungsten-filament valves, the operating costs may be divided under two headings:

(*a*) The cost of converting existing installations from tungsten-filament valves to thoriated-tungsten-filament valves.
(*b*) The savings brought about by reduced filament power consumption of the thoriated valve.

The capital cost of conversion varies widely, depending on whether the valve is a direct replacement both mechanically and electrically for the tungsten-filament valve. Valves CAT27, CAT30 and CAT29 are of this type, and the cost of converting a 100 kW h.f. transmitter is approximately £50. This is made up mainly of recalibration of meters and alteration to protective relay circuits. Where the valves are supplied from transformers, usually of the surge-limiting type, the cost is likely to be considerably increased. In the case of valve type 3Q/294E, which requires the fitting of a new valve jacket and various other mechanical alterations, the cost of converting a 100 kW h.f. transmitter is £400. The estimated savings in filament power of the thoriated valve compared with the tungsten-filament valve is tabulated in Section 4.2. However, in no case have the final average life figures for the thoriated valves been established, and the total savings have therefore still to be established. It may well be that these will be considerably increased as more information becomes available.

(4.1) Mechanical Aspects

The need for interchangeability in existing equipment of the thoriated-tungsten-filament valves has already been referred to, in order to reduce the capital cost of conversion. However, existing stocks of tungsten-filament valves have to be consumed, and, in addition, it is not usually possible to remove from service any of the equipment which is to be converted, except for a very short period. Various mechanical equipment used for handling valves, such as valve trucks, etc., should also be easily adaptable to handle the thoriated valves. There is thought to be a tendency for the thoriated valve to be more fragile than the tungsten equivalent, and the incidence of breakage in transit is being carefully watched to establish the need for any special packing or handling arrangements.

(4.2) Electrical Compatibility

When considering the conversion of existing tungsten-filament installations to thoriated-tungsten-filament valves, the source of filament power supply is important. In older installations, where motor generators are often used, no difficulty has been encountered in altering the output voltage and current over a wide range. However, voltage regulators and protective devices have required modification where the operating voltage of the thoriated valves have differed from that of the plain tungsten filament.

Savings in Filament Heating Power with Thoriated Tungsten Filaments compared with Pure Tungsten Filaments

CAT27	66%
3Q/294E	83%
CAT30	68%
3Q/260E	63%
BW165	69%
BR161	76%

In a typical installation of four 100 kW short-wave transmitters converted to thoriated tungsten filaments, the net power savings have been measured as 156 kVA on maximum demand per annum and 751 MWh per annum. This is equivalent to 13% reduction in maximum demand for the complete installation and 18% reduction in the total power consumption.

(5) CONCLUSION

The authors consider that there has now been sufficient experience in the use of valves with thoriated tungsten cathodes to recommend their extended use in all future equipment. It is possible that the full potentialities of thoriated tungsten cathodes have not yet been attained, and that, with improvements in vacuum technique and getters, the much greater emission which is available can be more fully exploited. This will result in more compact valve designs, for use particularly at the higher frequencies. It seems unlikely that, for the type of valve covered in the paper, the carburized thoriated tungsten cathode will be superseded by any other form of emitter for some years to come.

(6) ACKNOWLEDGMENTS

The authors wish to thank the Director of Engineering of the British Broadcasting Corporation for permission to publish the paper, and also to express their appreciation of the work of their colleagues throughout the Engineering Division of the British Broadcasting Corporation. They also wish to thank Standard Telephones and Cables, Ltd., the English Electric Valve Co., Ltd., the M.O. Valve Co., Ltd., and the Research Laboratories of the General Electric Co., Ltd., for permission to publish the information contained in the paper.

(7) REFERENCES

(1) LANGMUIR, I.: 'The Electron Emission from Thoriated Tungsten Filaments', *Physical Review*, 1923, **22**, p. 357.
(2) ANDREWS, M. R.: 'The Evaporation of Thorium from Tungsten', *ibid.*, 1927, **33**, p. 454.
(3) JENKINS, R. O., and TRODDEN, W.: 'Evaporation of Thorium from Carburized Thoriated-Tungsten Cathodes', *British Journal of Applied Physics*, 1959, **10**, p. 10.
(4) HORSTING, C. W.: 'Carbide Structures in Carburized Thoriated Tungsten Filaments', *Journal of Applied Physics*, 1947, **18**, p. 95.
(5) AYER, R. B.: 'Use of Thoriated Tungsten Filaments in High Power Transmitting Tubes', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 591.
(6) SCHWARSKOPF, P., and KIEFFER, R.: 'Refractory Hard Metals' (Macmillan, New York, 1953).
(7) SMITHELLS, C. J.: 'Tungsten' (Chapman and Hall, 1952).
(8) BECKER, K.: 'Die Kristallstruktur und der Lineare Warmeausdehnungskoeffizient der Wolframcarbide', *Zeitschrift für Physik*, 1928, **51**, p. 481.
(9) WORTHING, A. G.: 'The Thermal Expansion of Tungsten at Incandescent Temperatures', *Physical Review*, 1917, **10**, p. 638.
(10) AMERICAN SOCIETY OF METALS: 'Metals Reference Handbook' (Cleveland, 1948).
(11) WRIGHT, D. A.: Unpublished work.
(12) MAXWELL, J. C.: 'Electricity and Magnetism' (Oxford University Press, 1904), Vol. 1, Art 314.
(13) FORSYTHE, W. E., and WORTHING, A. G.: 'The Properties of Tungsten and the Characteristics of Tungsten Lamps', *Astrophysical Journal*, 1925, **61**, p. 172.
(14) ESPERSON, G. A., and ROGERS, J. W.: 'Studies on Grid Emission', *Transactions of the Institute of Radio Engineers*, 1956, **PGED 3**, p. 100.
(15) BELL, J., DAVIES, J. W., and GOSSLING, B. S.: 'High-Power Valves: Construction Testing and Operation', *Journal I.E.E.*, 1938, **83**, p. 176.

621.317.75

The Institution of Electrical Engineers
Paper No. 3165 M
Mar. 1960
©

AN IMPROVEMENT TO THE ELECTRON-TRAJECTORY TRACER

For use with the Resistance-Network Analogue

By J. VINE, M.Sc., and R. T. TAYLOR.

(*The paper was first received* 13*th June, and in revised form* 23*rd September*, 1959.)

SUMMARY

To increase the accuracy of the trajectory tracer improved methods of field determination on the resistance network have been investigated. One of these—an interpolation system applicable to axially symmetric fields—is described. This system has been constructed, and some results obtained with it for trajectories through an electron lens are given. These are compared with corresponding results obtained without the interpolation device and with results obtained by other methods.

It is seen that the interpolation system provides significantly increased accuracy for this particular type of problem, with little extra complication in the operation of the trajectory plotter.

LIST OF SYMBOLS

r, θ, z = Cylindrical polar co-ordinates.
V_B = Electrostatic potential at the network point B.
$\delta V_{BE} = V_B - V_E$ = Potential difference between network points B and E.
E_r, E_z = Electric field components.
E_{rK}, E_{zK} = Electric field components at the point K.
$\delta_r V_K, \delta_z V_K$ = Voltages proportional to E_{rK} and E_{zE}, respectively.
m = Integer expressing in meshes the radial co-ordinate of a network point.
ζ = Fractional part of radial co-ordinate of a point.
R_r = Resistance of interpolation control.
R_i = Input resistance of the trajectory tracer computing circuit.

(1) INTRODUCTION

The theory and construction of the electron-trajectory tracer have been described in a previous paper,[1] in which a number of projected alterations to the instrument were outlined. Shortly after the publication of that paper, work on the project was suspended, pending the investigation of alternative methods of trajectory tracing involving the use of an electronic digital computer. As a result of the investigations[2] the digital methods were deemed preferable, on balance, to the analogue method, and consequently it is unlikely that further development of the trajectory plotter will take place. Before the work ceased, however, one important improvement to the equipment was made, namely the incorporation of an interpolation system to give improved accuracy in the determination of the electric field, and the purpose of this paper is to complete the account of the trajectory plotter with a description of this.

(2) DETERMINATION OF THE ELECTRIC FIELD WITH A RESISTANCE NETWORK

It was shown in the previous paper that the main limitation to the accuracy of the trajectory plotter was the finite size of the resistance-network mesh. Apart from the truncation error incurred in any finite-interval method of integration, this introduces the disadvantage that the electric field components E_z and E_r (in the case of an axially symmetric problem, or E_x and E_y in the 2-dimensional case) are directly available at only a limited number of points.

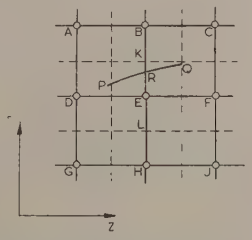


Fig. 1.—Determination of field component for a path element.

With reference to Fig. 1, where A, B, C, . . . H, J are network points and the dotted lines connect the mid-points of the meshes, consider the computation of the path element PRQ. The potential differences between neighbouring network points will be denoted by $\delta V_{BE} = V_B - V_E$, etc., where V_B is the potential at B, etc. Such a voltage may be regarded as proportional to the electric field component mid-way between the two points concerned. Voltages proportional to the z and r field components at a point, say K, will be denoted by $\delta_z V_K$ and $\delta_r V_K$, respectively, and, to avoid complication, proportionality constants will be ignored. Thus one can write

$$E_{rK} = \delta_r V_K = \delta V_{BE}$$

and

$$E_{zE} = \delta_z V_E = \tfrac{1}{2}\delta V_{FD}$$

the factor $\frac{1}{2}$ occurring because the potential difference δV_{FD} is measured across two meshes. Note that $\delta_z V_K$ is not expressible as a simple potential difference between two mesh points, although $\delta_r V_K$ is.

To compute the path element PRQ the trajectory plotter requires voltages $\delta_r V_R$ and $\delta_z V_R$, representing the field at the mid-interval point R. Neither of these is available in the above simple form and the computation method previously described would simply use $\delta_r V_K$ and $\delta_z V_E$, these being the nearest values so available. This method represents the field variation as a step function of r as shown in Fig. 2, where the continuous curves are the true variations in a typical case. Unless very large network models are used considerable errors can be introduced by the discrepancies between these curves, particularly in the region close to the axis, and a system of interpolation becomes necessary to provide a more continuous representation of the field. A simple example of such a system has been constructed and will now be described.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Vine is at the Research Laboratory of Siemens Edison Swan, Ltd., and Mr. Taylor is at the Research Laboratory of Associated Electrical Industries, Ltd.

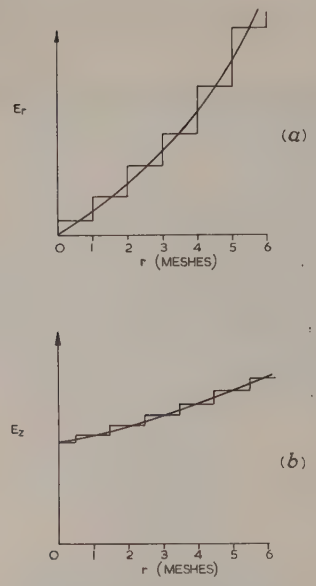


Fig. 2.—Variation of field components with r.

(a) Radial component, E_r.
(b) Axial component, E_z.
Curves show true variation in a typical case. Stepped lines show representation in the trajectory plotter without interpolation.

(3) DETERMINATION OF THE FIELD BY INTERPOLATION

The choice of unit trajectory intervals in the z direction, and the arrangement of the element relative to the mesh system as shown in Fig. 1, ensure the availability of field voltages at points having the correct z co-ordinate but displaced in the r direction from the mid-interval point (e.g. $\delta_r V_K$, $\delta_z V_E$ in the case of the element PRQ). Consequently, field variation with r must be considered but not variation with z, which, to a first approximation, cancels out over the element.

Confining attention to the axially symmetrical type of problem, it is well known that if second- and higher-order powers of r are ignored, one can write, for fixed z,

$$E_r \propto r$$
$$E_z = \text{constant}$$

It follows that, provided that network points are not too widely spaced, interpolation for E_z is unnecessary, while for E_r linear interpolation should suffice. These facts are evident from Fig. 2, where it can be seen that, close to the axis, errors in the field approximation approach 100% for E_r but are comparatively small for E_z. To define E_r along a mesh side such as BE (Fig. 1), therefore, a linear variation is assumed which is zero at $r = 0$ and equals δV_{BE} at r_K (co-ordinate of K). This representation is compared with the true curve in Fig. 3. It should be remarked that, although the absolute error in this representation may increase away from the axis, its percentage value does not necessarily do so, since E_r increases.

The expression for each of the straight-line segments in Fig. 3 is easily determined. With reference to Fig. 1, let the radial co-ordinates of points E, K and B be m, $m + \frac{1}{2}$ and $m + 1$ (mesh units), respectively. The voltage δV_{BE} may be designated $\delta_r V_{m+\frac{1}{2}}$ representing the radial field at $r = m + \frac{1}{2}$, and this, together with $E_r = 0$ at $r = 0$, gives

$$E_r = \frac{2r}{2m+1}\delta_r V_{m+\frac{1}{2}} \quad m < r < m+1 \quad . \quad . \quad (1)$$

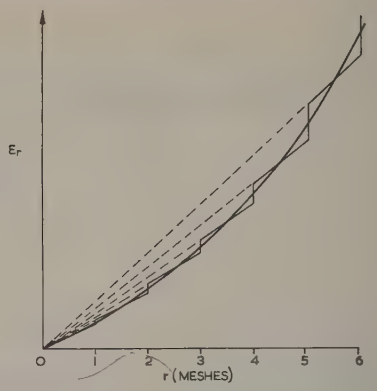


Fig. 3.—Representation of the radial field component by the interpolation system.

Note that this expression is dependent on only one network voltage. A superior interpolation system can be devised, assuming linear variation between two neighbouring field values such as those at K and L, but this involves the use of two mesh voltages leading to complication in the circuit necessary to perform the interpolation. For the purpose of demonstrating the possibilities of an interpolation system the present simpler system has been adopted.

Denoting the fractional part of r by ζ, the co-ordinate of R may be written $m + \zeta$, and eqn. (1) gives

$$\delta_r V_{m+\zeta} = \frac{2(m+\zeta)}{2m+1}\delta_r V_{m+\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (2)$$

A simple circuit which would give an output proportional to this quantity is shown in Fig. 4, where ζ is the fractional setting of the potentiometer.

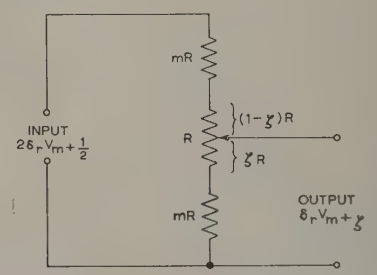


Fig. 4.—Principle of the interpolation circuit.

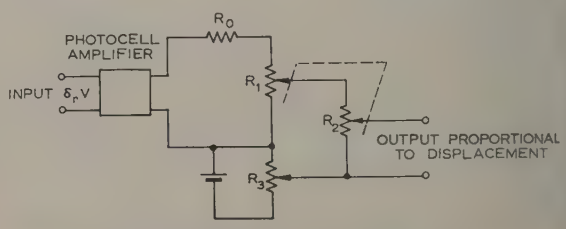


Fig. 5.—Basic circuit of the trajectory tracer.

(4) PRACTICAL INTERPOLATION ARRANGEMENT

Fig. 5 shows the basic computing circuit of the trajectory plotter, and the incorporation of the interpolation device into this requires consideration. The device of Fig. 4 could be

inserted either before or after the amplification stage, but in the first case a loading effect would disturb the network voltages, while in the second the interpolation circuit would itself be loaded by the input resistance of the computing circuit. To avoid the introduction of electronic buffer stages to eliminate these loading effects, the arrangement shown in Fig. 6 has been adopted. The

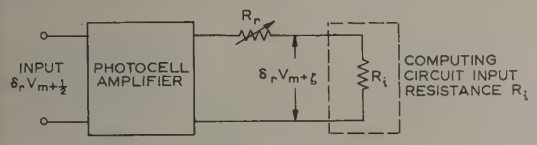


Fig. 6.—Practical arrangement of interpolation device.

variable resistor R_r is the interpolation control, in series with the input resistance R_i of the computing circuit. The gain of the amplifier and the factor 2 in eqn. (2) can be ignored as they simply change the proportionality constant relating the field to the network voltages, so that, essentially, a voltage $\delta_r V_{m+\frac{1}{2}}$ is applied to R_r and R_i in series, and it is necessary for the voltage $\delta_r V_{m+\zeta}$ to appear across R_i.

Therefore

$$\frac{R_r + R_i}{R_i} = \frac{\delta_r V_{m+\frac{1}{2}}}{\delta_r V_{m+\zeta}} = \frac{2m+1}{r}$$

so that

$$R_r = \left(\frac{2m+1}{r} - 1\right) R_i \quad . \quad . \quad . \quad . \quad (3)$$

The variation of R_r with r expressed by eqn. (3) is too complicated to be achieved with a continuously wound voltage divider and so this component is constructed with a switch and fixed resistors. To each switch position is assigned one value of r, the appropriate resistance to be connected being calculated from eqn. (3). Since the number of switch positions is limited, so is the possible number of interpolation points which must be suitably spaced across the range of r of interest. Since interpolation is most essential in the paraxial region the density of points should be highest near the axis.

To test the principle a system has been built covering just the first mesh, $m = 0$, with ten interpolation points at equal intervals, namely 0·05, 0·15, 0·25, . . . 0·95, and the circuit of this is shown in Fig. 7. The resistances are according to eqn. (3), with $R_i = 1\,250$ ohms.

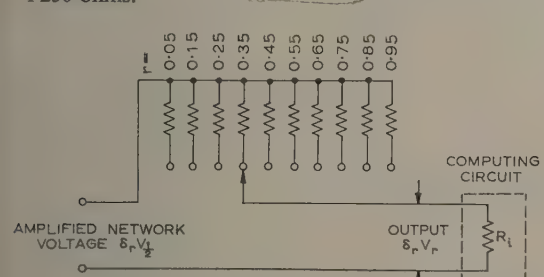


Fig. 7.—Interpolation circuit for the first mesh, $m = 0$.

(5) OPERATION OF THE TRAJECTORY PLOTTER

The operation of the plotter is complicated only by the necessity of setting the interpolation switch to the appropriate position at the start of each computation step. With reference to the element PRQ in Fig. 1, the position of P is known and an estimate is made of the r co-ordinate at R on the basis of the displacements obtained in preceding steps. The interpolation switch is then set to the position corresponding most closely with this estimate. Since the interpolation intervals are fairly large a rough estimate of the displacement usually suffices to determine the correct setting, but should this prove to be wrong according to the r-displacement subsequently obtained, the setting can be corrected and a new reading taken.

(6) RESULTS

To test the interpolation method, trial plots have been performed for Liebmann's[3] lens 5, which was set up on the network exactly as described in the previous paper.[1] Fig. 8 shows the

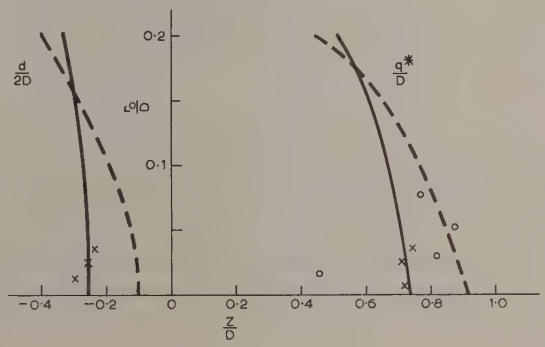


Fig. 8.—Comparison of results for lens 5.
—— Hand computation.
---- Liebmann's experimental results.
⊙ Results from trajectory tracer without interpolation.
× Results from trajectory tracer with interpolation.

results for three rays in comparison with curves obtained by alternative computation methods. The z co-ordinate of the focus, q^*/D, and that of the principal surface, $d/2D$, are plotted against the initial off-axis distance of the ray, r_0/D, D being the diameter of the lens. The broken curves are Liebmann's experimental results,[3] and some of the previously obtained trajectory-plotter results (without interpolation) are shown for comparison. It is seen that the interpolation procedure has considerably reduced the effect of finite mesh size, and gives results in good agreement with the computed curves. Fig. 9 shows two complete trajec-

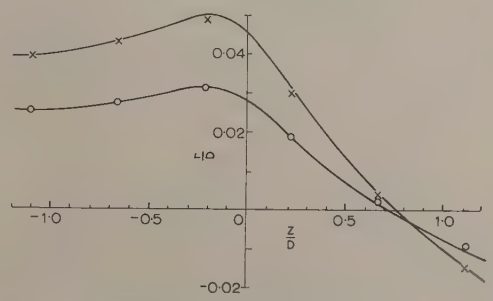


Fig. 9.—Trajectories through lens 5.
Curves are trajectory tracer results.
Points are results from integration of the ray-equation.
× Upper trajectory.
⊙ Lower trajectory.

tories obtained with the trajectory plotter together with points obtained by numerical computation.

Two different methods were employed for the check computations. The first, an integration of the ray equation similar to the method described by Liebmann,[4] was carried out with an

electronic digital computer. The second method was similar to that used in the trajectory plotter but employing a more careful interpolation procedure for both field components. Good agreement was obtained between the two methods as shown by Table 1.

Table 1

r_0/D	Plotter method		Ray equation integration	
	q^*/D	$d/2D$	q^*/D	$d/2D$
0			0·748	−0·253
0·05	0·713	−0·258	0·730	−0·261
0·1	0·687	−0·253	0·689	−0·252
0·15	0·621	−0·286	0·624	−0·275
0·2	0·524	−0·335	0·544	−0·294

The curves used in Fig. 8 represent an average of these two sets of results. The agreement of the different sets of computed results amongst themselves, and their disagreement with the experimental results of Liebmann, might indicate a discrepancy between the actual lens used in the experiments and its representation by the network model. This is not pursued further as the point of interest here is the comparison of the trajectory-plotter results with those obtained by other computational methods.

(7) CONCLUSIONS

The interpolation method described is simple, and gives a considerable improvement in the accuracy of the trajectory plotter at the expense of one extra operation per computation step. The trial plots were of paraxial rays, and it is probable that even greater accuracy would be obtained for trajectories further from the axis if the interpolation system were suitably extended. However, if such a device were to be developed further it might be advantageous to adopt a slightly more complicated method involving linear interpolation between two neighbouring field values. If applied for both field components, such a system would be generally applicable to problems on an *xy* network as well as on the axially symmetrical type.

(8) REFERENCES

(1) HAINE, M. E., and VINE, J.: 'An Electron-Trajectory Tracer for use with the Resistance Network Analogue', *Proceedings I.E.E.*, Paper No. 2800 M, December, 1958 (**106 B**, p. 517).

(2) VINE, J.: 'Application of a Combination of Analogue and Digital Computers to Electron Trajectory Tracing', *Computer Journal*, 1959, **2**, p. 137.

(3) LIEBMANN, G.: 'Measured Properties of Strong Unipotential Electron Lenses', *Proceedings of the Physical Society*, 1949, **62**, p. 213.

(4) LIEBMANN, G.: 'An Improved Method of Numerical Ray Tracing through Electron Lenses', *ibid.*, 1949, **62**, p. 753.

621.317.335

The Institution of Electrical Engineers
Paper No. 3159 M
Mar. 1960

IMPROVEMENTS IN THE PRECISION MEASUREMENT OF CAPACITANCE

By G. H. RAYNER, B.A., and L. H. FORD, M.Sc., Associate Members.

(*The paper was first received 7th July, and in revised form 16th September*, 1959.)

SUMMARY

All measurements of capacitance and resistance at the N.P.L. depend ultimately upon the Campbell standard of mutual inductance and upon frequency. Hitherto capacitance has been obtained from mutual inductance and resistance using a Carey–Foster type of network. The precision obtained with this in the Astbury bridge has been considerably improved since it was constructed 20 years ago, and an accuracy of repetition of about five parts in 10^6 is now obtained. Capacitance can also be derived from resistance and frequency; a Wien bridge has recently been constructed for this purpose. It has been designed to make a single precise measurement of capacitance (the repetition accuracy is two parts in 10^6) and thereby to complete a system of three measurements relating capacitance, resistance and mutual inductance in conjunction with frequency.

(1) INTRODUCTION

The measurement of capacitance at the National Physical Laboratory has been made in the past solely in terms of mutual inductance and resistance using some form of the Carey–Foster network, the present form being the Astbury bridge.[1, 2] An alternative method is to measure capacitance in terms of resistance and frequency. There are several networks which, in principle, may be used for this purpose: amongst these is the Wien bridge, and such a bridge has been recently designed and constructed at the N.P.L. As resistance is derived from mutual inductance and frequency by means of the Campbell–Ohm bridge, both methods can be reduced to the same basic form: namely, that capacitance is derived from mutual inductance and frequency. However, by using the two methods, a cycle of measurements is completed relating capacitance, resistance and mutual inductance in conjunction with frequency, and a check on the accuracy of the measurements is thereby obtained.

The measurement of capacitance involves resistance in both the Astbury and Wien bridges, and since capacitors are, in general, less stable than resistors or inductors, the Astbury and Wien bridges are best considered as alternative methods of measuring capacitance.

A precision Wien bridge has been described by Webb and Wood[3] for measuring capacitors of any value between 100 pF and 1 μF, a single capacitor being measured in terms of a single adjustable resistor and frequency.

Dunn[4] has recently adapted a commercial precision capacitance bridge for use as a Wien bridge for measuring a range of capacitances; like Webb and Wood he measured only one capacitor by means of one resistor.

A consideration of the Wien bridge circuit shows that there are disadvantages in restricting the measurement to one capacitor and one resistor of the network. It is necessary to determine the phase defects of the other two components, whereas if both the capacitors and both the resistors of the frequency-dependent arms are used in the capacitance determination, only the differences of the phase defects are required. The duplication of the capacitance and resistance measurements introduces an element of diversity, since two values of each are involved; nor does it take much extra time, since, in the single-capacitance method, the ratio of the two capacitances (or the two resistances) must also be measured.

Bridges for the simple comparison of capacitors or resistors are less exacting in their demands upon the design, construction and use than those, such as the Wien bridge, which relate dissimilar quantities, and, in general, the wider the range of a bridge the more difficult it becomes to obtain the highest precision; the bridge described below has therefore been designed to fulfil a single purpose with the highest precision. Its function is to calibrate one of the groups of the N.P.L. standard capacitors; a precision of observation of one part in 10^6 on capacitance was aimed at and the residual parameters have been determined to within the accuracy requisite for this precision. The bridge operates at a frequency of 1 kc/s; the resistors are fixed in value, while the capacitors can be varied over a small range to effect a precise balance. The values of the other standard capacitors are obtained by intercomparisons in a build-up process using a Schering bridge.

A detailed description of the Wien bridge is given since it forms a convenient unit that could be used in conjunction with a decade capacitor to establish the unit of capacitance wherever stable resistors and a frequency standard are available.

Since publishing the full description of the methods of calibrating capacitors on the Astbury bridge, an appreciable improvement in the accuracy has been obtained by refinements in the measurement techniques and in the determination of the residual parameters. A brief account is included of the way in which the improvements have been effected, so that the published description of this bridge may be brought up to date.

(2) WIEN BRIDGE

(2.1) Conditions for Balance

The circuit of the Wien bridge is shown in Fig. 1; the conditions to be satisfied at balance are given by the equations:

$$\omega^2 C_1 C_2 R_1 R_2 \left[1 + \frac{\gamma_1 - \gamma_2}{\omega C_1 R_1} + \frac{\tau_1 - \tau_2}{C_2 R_2} + \frac{R_4(\tau_3 - \tau_4)}{C_2 R_1 R_3} \right] = 1 \quad . \quad . \quad . \quad . \quad (1)$$

$$\frac{R_1}{R_2} + \frac{C_2}{C_1} + \frac{\gamma_1 + \gamma_2}{\omega C_1 R_2} - \frac{\tau_1 + \tau_2}{C_1 R_2} = \frac{R_4}{R_3} \quad . \quad . \quad . \quad (2)$$

where γ_1 and γ_2 are the power factors of C_1 and C_2, and τ_1, τ_2, τ_3 and τ_4 are the time-constants of R_1, R_2, R_3 and R_4. It is desirable that the ratio arms shall be approximately equal in order that any small difference in their time-constant may be eliminated by reversing them; this not only dispenses with the need to measure the time-constant difference, which would be required to an accuracy of within about 10^{-10} henry/ohm, but it has the advantage of introducing a second set of observations. If the ratio arms are equal eqn. (2) is satisfied for the first order when $R_1/R_2 = C_2/C_1 = 1/2$, and the values have been selected accordingly. For a frequency of 1 kc/s $R_1 = 199$ ohms, $R_2 = 398$ ohms, $C_1 = 0{\cdot}7998\,\mu$F and $C_2 = 0{\cdot}3999\,\mu$F; C_1 and C_2 are adjusted to obtain a balance by variable air capacitors, which give a range of about $\pm 0{\cdot}4\%$ on the above values.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the National Physical Laboratory.

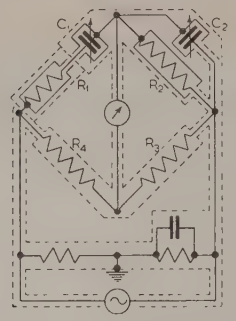


Fig. 1.—Wien bridge.

With these values for the resistors and capacitors the products $\omega C_1 R_1$ and $\omega C_2 R_2$ are both approximately equal to unity and, when the term $\tau_3 - \tau_4$ is effectively eliminated by reversing the ratios, the working equation relating capacitance, resistance and frequency reduces to

$$\omega^2 C_1 C_2 R_1 R_2[1 + (\gamma_1 - \gamma_2) + \omega(\tau_1 - \tau_2)] = 1 \quad . \quad (3)$$

It will be seen that it is only necessary to determine the differences of the power factors and time-constants. As mentioned above, this is an advantage since considerably greater accuracy can be attained in measuring the differences of these quantities than in measuring their values absolutely.

(2.2) The Practical Circuit

The elimination of certain residual effects presented some problems in the arrangement of the bridge components which had no strict solution, but a satisfactory compromise has been sought and adopted. The difficulties arise largely from the need to be able to compare the bridge components in arms 1 and 2 with the appropriate working standards of the N.P.L.; for these comparisons, current and potential terminals for R_1 and R_2 must be accessible, together with screened terminals for C_1 and C_2.

Provision must also be made for isolating at least one terminal of the components, but, at the same time, it is necessary to maintain the screening arrangements as indicated in Fig. 1.

Fig. 2 shows diagrammatically the arrangement which has been adopted; the components are connected by inserting plugs to make connection between the appropriate blocks. The blocks

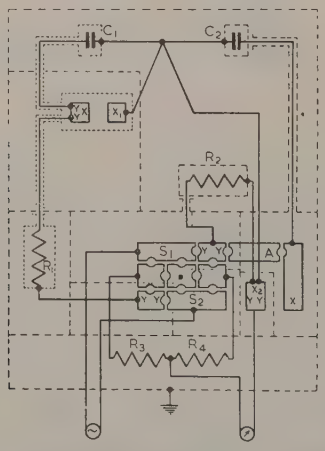


Fig. 2.—Wien bridge. Diagram of connections.
X. Connection for comparison with external capacitors.
YY. Connections for 4-terminal resistance measurements.

to which it is necessary to make external connection are indicated in Fig. 2; those marked X are used when comparing C_1 and C_2 with the standard capacitors, and those marked Y are for the measurement of R_1 and R_2 as 4-terminal resistors.

The plugs that must be inserted for the various measurements are shown in Fig. 3. Fig. 3(*a*) shows the plug connections for the Wien bridge and Fig. 3(*b*) the change required to reverse the

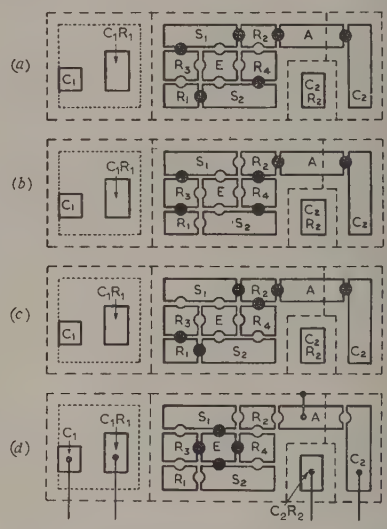


Fig. 3.—Terminal and plug blocks (to scale).
---- Earth screens.

ratio arms R_3 and R_4. It will be seen that this reversal is achieved by transferring the connection to the block marked R_1 from R_4 and S_2 to S_1 and R_3, and vice versa for the block R_2; the reversal of the ratios can also be effected by the arrangement shown in Fig. 3(*c*). In each of these changes the bridge supply points are transferred from one block to another, but as it has been found that the same result is obtained whichever way the ratio arms are reversed it is concluded that no error is introduced.

Fig. 3(*d*) shows how the plugs are inserted for the comparison of C_1 and C_2 with the standard capacitors; it will be seen that the four blocks, S_1, S_2, R_3 and R_4, are connected to the central earthed block; this, together with screening underneath the blocks, reduces the direct capacitance between the blocks R_1 and R_2 to less than 0·1 pF. In addition, the connecting block A is earthed, thereby eliminating stray capacitance between the blocks R_2 and C_2. For comparison of C_1 with the standard capacitor, screened leads are connected to the blocks marked C_1 and C_1R_1, and for the comparison of C_2 the leads are connected to the blocks C_2 and C_2R_2.

The leads from the junction of C_1 and C_2 to the connecting blocks C_1 and C_2R_2 (X_1 and X_2 in Fig. 2) have been made identical, and it can be shown that in this case their resistance and inductance introduce no errors and that the capacitances which are operative in the Wien bridge and their respective power factors are accurately compared. These measurements are made by substitution using a Schering bridge; corrections of three and two parts in 10^6 have to be applied for the series inductance and series resistance, respectively, of the leads connecting the standard capacitors and the Wien bridge.

In a similar manner R_1 and R_2, which are compared with standard resistors of similar value, are precisely those required for the Wien bridge; these resistance comparisons are made with

direct current. The resistors are made of fine wire and their change in resistance due to eddy currents (skin effect) from d.c. to 1 kc/s is negligible.

The measurement of the difference in their time-constants presents some difficulty because of the separation of the terminals. However, two different methods of connection to the resistors gave the same value for the difference, namely $\tau_1 - \tau_2 = -0{\cdot}0011\,\mu$H/ohm; thus the correcting term $\omega(\tau_1 - \tau_2)$ is -7×10^{-6} and is known to within one part in 10^{-6}. The method of making this measurement has been described in a recent paper.[5]

Undesirable stray capacitances between the various bridge components have been eliminated by earth screens, and the only capacitance that requires further consideration is that to earth from the lead between C_1 and R_1 where it passes through the earth screen separating these two components. This earth capacitance, C_E, was measured during the assembly of the bridge and is 0·5 pF; its effect is to increase the series resistance of arm 1 by a fraction C_E/C_1, namely 0·6 part in 10^6, which corresponds to 0·3 part in 10^6 on the unit of capacitance and is negligible.

(2.3) Layout of Bridge Components

The general arrangement of the bridge and the disposition of the components are shown to scale in Fig. 4. In the front, and

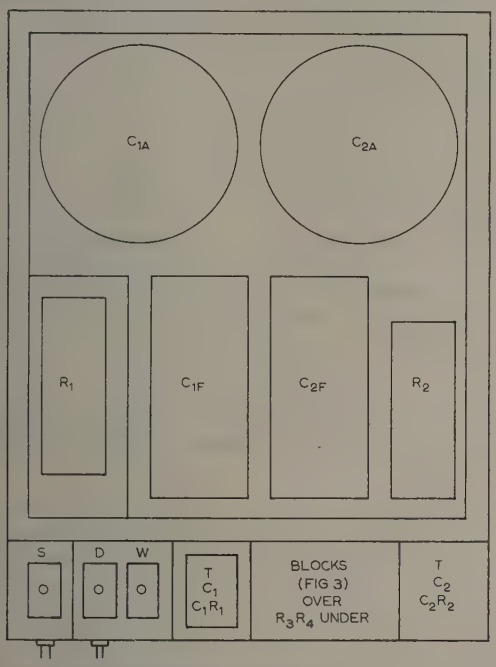


Fig. 4.—Layout of components (to scale).
S, D, W. Source, detector and Wagner switches.
T. Terminals.
Overall dimensions: 14 in × 11 in × 8 in high.

accessible through a hinged panel of the top cover, are the plug blocks and terminals. It will be seen that the latter, including terminals C_1 and C_1R_1, are near the edge of the box. There are holes in the front panel to admit the leads for making the comparisons of the bridge resistors and capacitors with the external standards, and the terminals of the capacitors register with those of the comparison standards. Adjacent to the terminals are switches (not shown in Fig. 2) for reversing the supply and detector connections, and for joining one of the detector points to earth when balancing the Wagner earthing arms; these are separate from the bridge, and, as they follow the normal practice, have not been included in the circuit diagram.

The fixed and adjustable capacitors, $C_{1F}C_{1A}$ and $C_{2F}C_{2A}$, are the bulkiest components of the bridge. The fixed capacitors are mica units which are mounted in sealed cases; they were kept in a desiccator for about four years before being assembled in the bridge. These capacitors have been put as close as possible to the connecting blocks in order to reduce the lengths of the leads, and thus to reduce the series resistance and inductance. The variable capacitors for C_1 and C_2 have ranges of about 600 and 300 pF, respectively, and can be read to within one part in 10^3.

The resistors R_1 and R_2 are wound with No. 44 s.w.g. Evanohm wire (an alloy of nickel and chromium, with small amounts of aluminium, copper and iron) on high-quality mica; each is mounted in its individual screening case, and that containing R_1 is completely enclosed in an earthed screen. The resistors are placed beside the capacitors and near the terminals so that the leads are also short.

The resistors for the ratio arms are wound with the same wire and are of 200 ohms each; they are wound on a single mica sheet with a centre tap and are in a screened compartment under the terminal blocks.

The bridge components, plug blocks and switches are mounted on Perspex panels which are fixed to a frame of brass. This frame carries the internal earthed screens and the external panels which are all independently detachable so that access can be had to the components and connections.

(2.4) Supply and Detector

The 1 kc/s supply to the bridge is obtained direct from the laboratory frequency standard, and to reduce harmonics it is fed through a tuned filter to a screened isolating transformer.

A narrow-band selective amplifier[6] followed by an oscilloscope is used as the detector, the connection from the bridge being through a screened transformer. The sensitivity of the system is such that 1 volt applied to the bridge suffices to give an observational accuracy of one part in 10^6; with this voltage the changes in the values of the resistors because of self-heating are negligible. The harmonic voltages at the input to the amplifier are sufficiently small for the intermodulation products to be entirely negligible.

(3) DERIVATION OF RESISTANCE FROM MUTUAL INDUCTANCE AND FREQUENCY

Since both the Astbury and Wien bridges require a knowledge of resistances for the determination of capacitance, it is appropriate to indicate in outline how resistance is derived from the primary mutual inductance standard.[7,8] Resistance is measured in terms of mutual inductance and frequency by means of the Campbell–Ohm bridge, the basic circuit of which is shown in Fig. 5.

The circuit contains two resistors, r and R, and two mutual inductors, M_1 and M_2. These have the same nominal value

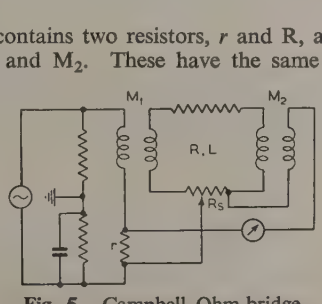


Fig. 5.—Campbell–Ohm bridge.

(10 mH) as the primary standard and can be compared with it; r is a 4-terminal 1-ohm resistor and R is the total resistance (about 61·8 ohms) of the closed secondary mesh. The supply to the bridge is at 125 c/s and is derived from the frequency standard. The conditions to be satisfied at balance are as follows:

$$\omega^2 M_1 M_2 + Rr = 0 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$M_1 R_s = Lr \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

If r and R are measured in terms of the resistance unit which is maintained by the standard 1-ohm resistors of the N.P.L., the difference between $\omega\sqrt{(M_1 M_2)}$ and $\sqrt{(rR)}$ shows by how much the N.P.L. unit of resistance differs from the absolute unit; determinations in 1936 and 1951 gave a difference of $+2 \times 10^{-6}$, with an estimated limit of uncertainty of $\pm 15 \times 10^{-6}$.

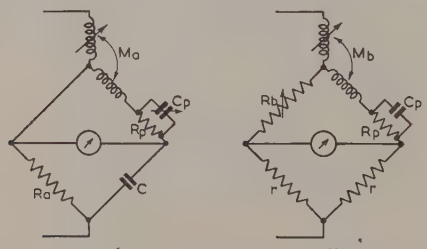
A number of small correcting terms have to be applied to the simple balance equations given above, but they are known and all the observations on the bridge are made to an accuracy of within at least one part in 10^6; this is also the repetition accuracy of the bridge. There may also be unsuspected systematic errors of a few parts in 10^6, and the overall limit of error of the measurement is estimated to be about five parts in 10^6.

An additional uncertainty, which is common to all the measurements of impedance, arises from possible differences between the value calculated for the primary standard and that realized in practice; this error is estimated not to exceed ten parts in 10^6, and gives the total uncertainty of 15 parts in 10^6. Errors in the primary standard do not, of course, affect the agreement which should be obtained in the capacitance measurements on the Astbury and Wien bridges.

(4) ASTBURY BRIDGE

Full accounts of the Astbury bridge[1] and the methods of maintaining the unit of capacitance[2] at the N.P.L. have already been published, and it will suffice to give a brief résumé, indicating where improvements have been made.

The Astbury bridge comprises two networks, one being a Carey–Foster and the other a Campbell–Heaviside network, as shown in Figs. 6(*a*) and (*b*). The former contains the capacitor,



(*a*) (*b*)

Fig. 6.—Astbury bridge (screens omitted).
(*a*) Carey–Foster network.
(*b*) Campbell–Heaviside network.

C, under test and it is balanced first by adjusting M and C_p. The circuit is then changed to that of Fig. 6(*b*); this is an equal-ratio bridge, and the effective resistance of the series arm—which is identical with that of Fig. 6(*a*)—is balanced and measured by the resistor R_b. These measurements are usually carried out at 1 kc/s. It can readily be shown that

$$C = M_a[1 + (\sigma_a - 2\sigma_b)/R_b]/R_a R_b \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\gamma = \omega[(2M_b - M_a)/R_b + \tau_a + \tau_b] + \sigma_a/\omega M_a \quad . \quad (7)$$

where γ is the power factor of the capacitor C, σ is the impurity of the mutual inductor, and τ is the time-constant of the adjustable resistor, which is used for both R_a and R_b. The values R_a and R_b are always chosen so that they can be compared with standard resistors of the same nominal value, and the resistor can be connected to a d.c. bridge for this comparison. The mutual inductor[9] is calibrated at its 10 mH setting by comparison at 10 c/s with the primary standard, and the values at the other settings are obtained by a build-up process.

The values required in eqn. (6) are those applicable at 1 kc/s, and it is therefore necessary to make corrections for the changes with frequency of the mutual inductor between 10 and 1 000 c/s and of the resistor between 0 and 1 kc/s. Also the difference $(\sigma_a - 2\sigma_b)$ of the impurities must be determined at 1 kc/s. The frequency change of the mutual inductor was the only one of these that was significant for the accuracy needed in capacitance measurements when the bridge was constructed (i.e. a few parts in 10^5) and it was sufficient to show that the others were insignificant. The value of the impurity, it may be noted, was reduced to 0·002 ohm at 1 kc/s by means of a compensating circuit.

Owing to the greater accuracy of which the Wien bridge is capable it was necessary to determine the smaller as well as the larger correcting terms with an accuracy that corresponds to a precision approaching one part in 10^6 on capacitance.

The frequency coefficient of the bridge as a unit has been determined from measurements made over a range of frequencies on air capacitors and on mica capacitors of known frequency change. The change in resistance with frequency has been determined at all the settings normally used for R_a and R_b by comparison with simple resistors which are wound with fine wire, so that the resistance change due to skin effect is negligible in the a.f. range. The resistance was found to decrease between 0 and 1 kc/s by amounts varying from one part in 10^6 at the 50-ohm setting to 27 parts in 10^6 at 2 kilohms. The impurity of the mutual inductor cannot be measured directly under the conditions of use in the Astbury bridge, but if all the other quantities in eqn. (7) are known with sufficient accuracy this equation can be used to obtain σ_a, the impurity at the setting M_a of the inductor. Since, by the choice of suitable values for C, R_a and R_b, eqn. (6) can be satisfied and the bridge balanced for any value of M_a, the impurity may be obtained for any setting of the mutual inductor. In this way the impurity has been obtained for the settings of M_a and M_b normally used, and the term $\sigma_a - 2\sigma_b$ is estimated to be known to about 0·000 2 ohm, which gives a maximum uncertainty of one part in 10^6 on capacitance for the smallest value of R_b normally used, namely 200 ohms.

The effects of stray mutual inductances have also been examined; a mutual inductance between the inductor and the rest of the circuit affects the measurements by one part in 10^6. Another mutual inductance in the bridge wiring adds a small amount to the value of the mutual inductor; this produces a mean overall correction of four parts in 10^6 in the calibration of the standard capacitors.

Since the previous account of the calibration of capacitors was published,[2] of the above correcting factors only the revision of the frequency coefficients of the bridge and of the resistor have made changes of more than ten parts in 10^6; these were remeasured in 1955, and, when combined, they reduce the capacitances assigned to the standard capacitors by 51 parts in 10^6. The other factors taken together add two parts in 10^6 to the values.

(5) EXPERIMENTAL PROCEDURE

The standard capacitors contain mica units, ranging from 0·01 μF up to a total of 1 μF, which are accurately additive and provide a precise scale of capacitances, and the following procedure has been used in recent years for their calibration. Units

having three capacitances, usually 0·01, 0·1 and 1 μF, are measured on the Astbury bridge, using two resistance settings for each; thus giving six measurements in all. The units are then intercompared on a Schering bridge, and, giving approximately equal weight to each of the Astbury bridge measurements, the most probable self-consistent values are assigned to the various units. The Schering bridge measurements also provide accurate values for the difference of power factor between the units; the power factors themselves are derived from a low-loss 3-terminal air capacitor.

To compare the unit as derived from the Astbury bridge with that from the Wien bridge, the capacitors giving nominal total values of 0·8 and 0·4 μF must be compared with the capacitors in the Wien bridge. The resistors of the Wien bridge have to be compared with the d.c. standards and observations on the bridge itself have to be made.

In order to eliminate the possibility of changes in the components of the bridges and in the standard capacitors, ideally all the measurements should be made simultaneously; this is clearly impossible, and even with two observers making measurements concurrently, two hours are required to complete the observations. There is evidence which suggests that the standard capacitors are the most likely to cause errors due to variations in their values during the sequence of measurements. The units of these capacitors have temperature coefficients of between +10 and +20 parts in 10^6 per deg C, and although the room is thermostatically controlled there are minor temperature fluctuations and gradients that may give variations of a few tenths of 1° C. It has been observed that, when the Schering bridge measurements are made before as well as after those on the Astbury bridge, there may be small differences of one or two parts in 10^6 in the relative values of some of the units.

(6) RESULTS AND CONCLUSIONS

Six simultaneous determinations of capacitance on the Astbury and Wien bridges were carried out over a period of nine months. Table 1 gives the difference, $C_A - C_W$, where C_A and C_W are the capacitances determined by the two bridges.

Table 1

DIFFERENCE OF ASTBURY AND WIEN BRIDGE CAPACITANCES

Parts in 10^6

$C_A - C_W$		+1	+1	0	−12	−1	−3	Mean −2
Astbury bridge deviations	Highest	+13	+5	+6	+11	+10	+13	
	Lowest	−10	−6	−10	−12	−9	−11	
	Mean	8	4	5	7	6	6	

In addition, to illustrate the self-consistency of the Astbury bridge measurements, for each capacitance determination there are recorded the deviations of the highest and lowest values from the assigned value for the set of six Astbury bridge measurements, together with the mean magnitude of the deviations of all six measurements.

The mean difference of the Astbury and Wien bridge capacitances for the six sets of measurements is two parts in 10^6, which is entirely negligible from all points of view, since of the many scores of observations that are involved in any one comparison only a few are made with an accuracy better than one part in 10^6, and also since, as has been pointed out above, the standard capacitors may sometimes alter by this amount in the course of the series of measurements. It will be seen that the fourth measurement gave a result differing by ten parts in 10^6 from the mean, whereas all the others were within four parts in 10^6 of each other. However, it is considered that the self-consistency of these five, rather than the divergence of the one, is the more remarkable.

After careful appraisal of the two methods we are of the opinion that the Wien bridge is likely to be the more consistent and accurate; the bridge has been specifically designed for a single measurement of the highest precision. The Astbury bridge, on the other hand, is capable not only of measuring capacitance over a wide range, but of calibrating self-inductors as well; the possibility of its use to measure capacitance to an accuracy approaching a few parts in 10^6 was not envisaged when it was made, 20 years ago.

It is concluded from the results that the unit of capacitance is consistent with the units of resistance and inductance, as maintained at the N.P.L., to within ten parts in 10^6, and, making allowance for the uncertainty in the primary standard of mutual inductance, that the absolute value of the unit is correct to within 20 parts in 10^6.

A comparison of these figures with the stability of the standard capacitors may be of interest. The most stable unit of these (0·2 μF in value) has shown fluctuations over a range of about 30 parts in 10^6 during the last five years.

An analysis of the behaviour of all the precision capacitor units of 0·1 μF and above at the N.P.L. shows that, in the same period, the average fluctuation is about 60 parts in 10^6 and that the greatest is 120 parts in 10^6. The smaller units (less than 0·1 μF) have, on the whole, shown considerably larger variations ranging from 100 parts in 10^6, for a unit manufactured 25 years ago, to 350 parts in 10^6, for some 0·01 μF units made nearly ten years ago; these are still continuing to increase in capacitance at the rate of about 60 parts in 10^6 per year.

The capacitors are units of commercial manufacture and our experience shows that their performance is representative of precision capacitors maintained under good conditions. It will thus be seen that the accuracy of capacitance measurement at the N.P.L. is appreciably better than can be used in practice owing to the instability of even the best standard mica capacitors.

(7) ACKNOWLEDGMENT

The work described has been carried out as part of the research programme at the National Physical Laboratory, and the paper is published by permission of the Director of the Laboratory.

(8) REFERENCES

(1) ASTBURY, N. F., and FORD, L. H.: 'The Precision Measurement of Capacitance', *Proceedings of the Physical Society*, 1939, **51**, p. 37.

(2) RAYNER, G. H., and FORD, L. H.: 'The Calibration of Capacitors of the National Physical Laboratory', *Journal I.E.E.*, 1948, **95**, Part II, p. 312.

(3) WEBB, J. K., and WOOD, H. B.: 'The Precise Measurement of Capacitance', *Proceedings I.E.E.*, Monograph No. 100 M, May, 1954 (**102** C, p. 3).

(4) DUNN, A. F.: 'Determination of the Unit of Capacitance', *Canadian Journal of Physics*, 1959, **37**, p. 35.

(5) RAYNER, G. H., and FORD, L. H.: 'The A.C. Properties of Resistors and Potential Dividers at Power and Audio Frequencies, and their Measurement', *Journal of Scientific Instruments*, 1957, **34**, p. 190.

(6) RAYNER, G. H.: 'A Selective Detector Amplifier for 10–10 000 c/s', *ibid.*, 1953, **30**, p. 17.

(7) HARTSHORN, L., and ASTBURY, N. F.: 'The Absolute Measurement of Resistance by the Method of Albert Campbell', *Philosophical Transactions of the Royal Society*, A, 1937, **236**, p. 423.

(8) RAYNER, G. H.: 'An Absolute Measurement of Resistance by Albert Campbell's Bridge Method', *Proceedings I.E.E.*, April, 1954 (**101**, Part IV, p. 250).

(9) ASTBURY, N. F., and FORD, L. H.: 'A Screened Sub-Standard Inductometer', *Philosophical Magazine*, 1938, **25**, p. 1009.

621.374.32 : 681.142

The Institution of Electrical Engineers
Paper No. 3167 M
Mar. 1960
©

FERRORESONANT SYSTEMS OF CIRCUIT LOGIC

By Professor J. G. SANTESMASES, M. ALIQUE, Dr.Sc., and J. L. LLORET, Bs.Sc.

(*The paper was first received* 15*th July, and in revised form* 28*th September*, 1959.)

SUMMARY

The logical behaviour of a basic ferroresonant circuit is described. A procedure for obtaining the 'and' and 'or' fundamental circuits through ferroresonant elements is reported. The possibility of coupling these fundamental circuits allows the development of a ferroresonant system of circuit logic. A ferroresonant inhibit circuit as well as the simplification it permits in the logical system are studied. The results obtained are described.

(1) INTRODUCTION

In a *RLC* circuit connected to an a.c. source, where any of the parameters is a function of current, the current/voltage curve for the whole circuit may be a multi-valued function. Thus, the presence of a ferromagnetic inductor in a circuit results in certain phenomena, some of which are known as ferroresonance.

The simplest circuit where ferroresonance may take place consists of a capacitor and a ferromagnetic-core inductor connected in series with an alternating voltage, or in parallel with an alternating constant-current generator. In either case, more than one stable operation condition will occur under certain circumstances.

The earliest papers on the problems arising from the presence of a ferromagnetic inductor in a resonant circuit were published at the beginning of this century; phenomena that were first considered in connection with radiotelegraph stations became noticeable in power distribution networks. The problem was studied by several research workers.[1-5] Later, interest on the subject decreased for some time, although its study was continued and papers were published from time to time; new applications were found, such as frequency multipliers, voltage stabilizers, automatic control systems without moving contacts, etc.

The first papers on bistable ferroresonant circuits applicable to digital computers were published in 1952. Isborn[6] described a series ferroresonant flip-flop circuit, and one of the present writers[7, 8] described a parallel ferroresonant flip-flop circuit with d.c. polarized cores.

So employed, ferroresonant circuits show many advantages over other circuits employing thermionic valves, such as greater reliability, longer life, lower power requirements, continuously available read-out and outputs of elements showing enough energy to allow the data to shift either to one element or simultaneously to more. Therefore the study of ferroresonance and its applications is highly interesting, and lately several works on this subject have been published.[9-18]

Between them, only three papers deal with the application of ferroresonance to combinational or decision circuits. For the 'and' and 'or' circuits Cutler[12] approaches the logical operation by means of a diode gate before the trigger circuit. For the inverter circuit he suggests two opposite-direction trigger windings. Arbon and Jones[14] describe these circuits and study the possibility of employing cores with polarization. Proebster[17] has carried out flip-flop-circuit interconnections and achieves logical operations through intermediate static magnetic cores.

As a common characteristic of these papers we find that ferroresonant elements are employed only as intermediate memory elements, supplying energy, if necessary to the diode or static-core circuits, which are those which actually carry out the logical operation.

In the paper we propose systems of circuit logic in which logical operation is achieved by the special connection of inductors employed in constituting the ferroresonant elements. They are those which we can actually designate by ferroresonant logical circuits.

(2) BASIC FERRORESONANT ELEMENT

If a ferroresonant circuit, Fig. 1, is connected to an a.c. source with low output impedance, we know[5] that a voltage/current

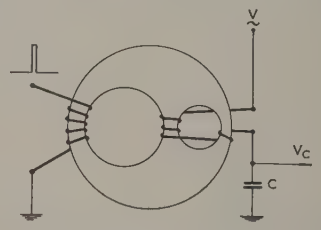


Fig. 1.—Ferroresonant element showing the relative winding positions.

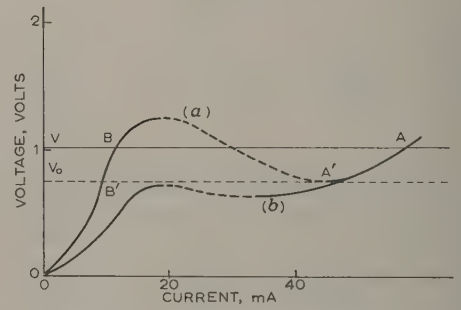


Fig. 2.—Ferroresonant characteristics.

(r.m.s.) ferroresonant curve for the element [Fig. 2, curve (*a*)] can be obtained when a suitable capacitance is used. If the a.c. source voltage stays at a value *V*, the state of the element can correspond to that of one of the points A and B, the intersection with the negative slope branch being unstable. Suppose that the ferroresonant element is in the B state, i.e. of low current; if a current pulse is sent to the core control winding, the characteristic will be momentarily modified to approach curve (*b*).[9] Therefore, A will be the only possible working-point, as in this zone (*b*) does not differ relevantly from (*a*). At the end of the pulse in the control winding, and once the charac-

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Prof. Santesmases is Director of, and Señores Alique and Lloret are at, the Instituto de Electricidad y Automática, Madrid, Spain.

teristic has recovered its form (*a*), the ferroresonant element will remain at the A state, i.e. high current, since it remains a stable operating-point. It will immediately be observed that in such a ferroresonant system it will not be possible to return from point A to B by acting on the control winding.

However, if during an interval the alternating voltage is lowered to a value V_0 below the minimum A′ of curve (*a*), all points of type A cease to exist, and the ferroresonant element jumps to B′, returning to point B when the generator recovers its normal voltage. Therefore the jump to the 'low' state may be obtained by means of a pulse modulating the amplitude of the alternating voltage.

On the other hand, it is found experimentally that, by keeping the generator voltage steady, the voltage, V_c, through the ferroresonant capacitor varies according to the frequency, as shown on Fig. 3. For a frequency f, the system may show both stable

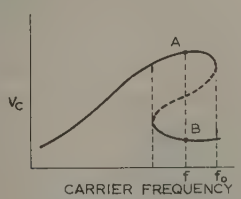


Fig. 3.—Alternating voltage through the ferroresonant capacitor as a function of the carrier frequency.

states A and B; therefore, from point A (high) it will be possible to jump to the state B (low) by means of a frequency-modulating pulse which raises f to a value higher than f_0.

A two-hole core, of the transfluxor type, has been chosen, mainly because of the very small voltage induced by the a.c. winding on the control of the core. We can therefore dispense with the usual systems aimed to avoid this induced voltage in the control winding, such as the connection of two cores with opposed windings.

From the logical standpoint, the basic ferroresonant element just described constitutes an *R–S* flip-flop circuit. The input *S*—the set line which corresponds to the core control winding—puts the element in the '1' state (high), and input *R*—the reset line—puts it into the '0' state (low), exciting a modulation in frequency or amplitude of the generator voltage. The system will remain in the '1' state from the time a set pulse reaches the control winding until a reset pulse excites the appropriate carrier modulation.

However, we can time the arrival of the reset pulses periodically by constituting a clock sequence which defines intervals between two consecutive pulses in which the system will remain at '0' or jump to '1' according as a pulse does or does not reach the control winding. Therefore, the ferroresonant element behaves in such a way that output depends only on what happens at the input within this same interval. This behaviour fits in with the so-called decision or combinational circuits.

On the other hand, if we add to the previous circuit a network storing that information existing in the output (for example a demodulator) we can make the output useful in the following interval between clock pulses. In the $(n + 1)$th interval the information in the intermediate storage element will be the same as that reaching the ferroresonant element in the nth interval. As a whole the system will then behave as a delay flip-flop circuit which may be the starting-point in achieving a shift register.

(3) DECISION OR COMBINATIONAL CIRCUITS

To begin with we shall study the combinational circuits from the basic ferroresonant element; this is entirely feasible because when different cores are connected through their a.c. windings in series-parallel assemblies the characteristic of the complex (or compound) inductor is adequate to achieve ferroresonance when it is connected to a capacitor, such as was shown in the description of the basic ferroresonant element. The resulting complex ferroresonant element has as inputs the control windings of each of the cores of which it is made up, and has an output which can, for example, be taken from the capacitor. An important advantage of this arrangement is that, owing to the arrangement of the cores, inputs are completely independent of each other. As a matter of fact, as the two windings of any core are not coupled, the connection of the alternating windings of two cores does not establish any link whatsoever between their control windings.

We shall next describe the setting-up of the 'and' and 'or' fundamental logical circuits, and we shall then proceed to those cases represented by the association of these circuits which result in the formation of rather more complex ferro-resonant combinational circuits.

(3.1) 'And' Ferroresonant Circuit

It is reasonable to admit, and we have experimentally verified, that the voltage/current characteristic of a set of equal non-linear inductors, arranged in series, may be determined, starting from the individual characteristics, by operating in the following manner: if all voltages corresponding to each current in the individual characteristics are added, their sum will be the

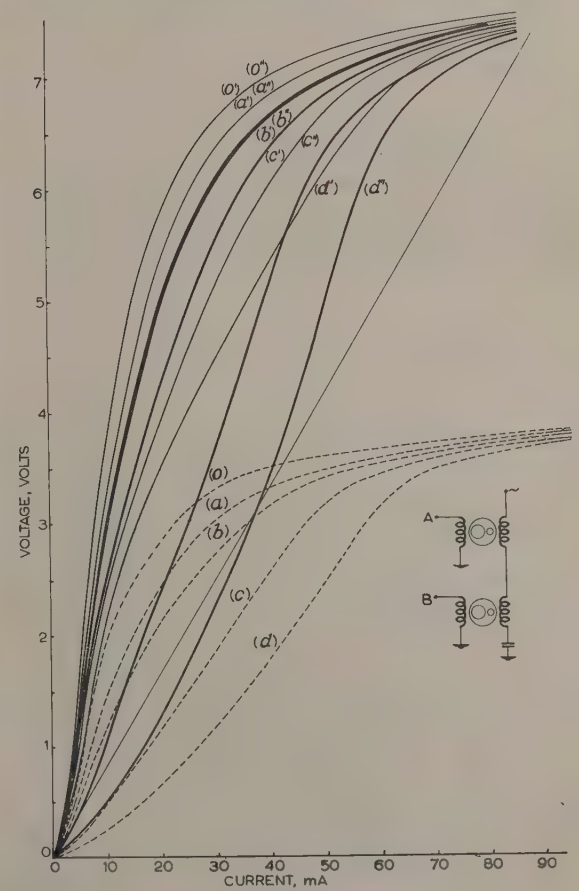


Fig. 4.—Voltage/current characteristic of a set of two equal inductors coupled in series.

voltage corresponding to that value of current in the characteristics of the set. Therefore, the knee of the characteristic of the set has the same current as, and its voltage co-ordinate is n times larger than, those corresponding to the knee of the characteristic of each one of the inductors in the set.

For instance, consider two equal inductors in series. Such an example is by no means restrictive. In Fig. 4 the curve (o') corresponds to the 2-inductor system and curve (o) to just one of them; the dotted lines represent the characteristics of each inductor with polarization. The curves (a')–(d') have been drawn by adding curve (o) to curves (a)–(d), i.e. they correspond to polarization being applied on only one of the inductors. And finally, if we suppose equal polarization on both inductors, curves (a'')–(d'') are obtained.

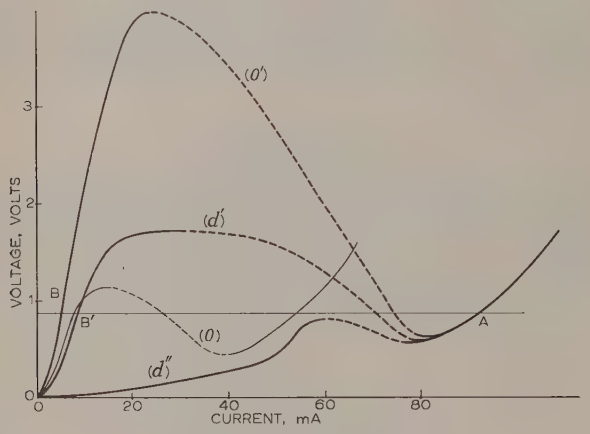


Fig. 5.—Ferroresonant curves of a system composed of two inductors in series and a capacitor.

Fig. 5 shows the ferroresonant characteristics corresponding to the system shown to the right of Fig. 4. Only the curves corresponding to (o), (o'), (d') and (d'') in Fig. 4 are shown. Suppose that the generator voltage remains steady at V; it can be seen that the system jumps from the 'low' to 'high' state only in the case of curve (d''), i.e. with both inductors polarized, and that curves such as (d') will never achieve the jump to the 'high' state, because, in any case, their maximum will be above the maximum corresponding to the reference curve (o).

We could apply the same results, with identical reasoning, to a set of n equal inductors. A set of n equal non-linear inductors in series with a suitable capacitor constitutes a ferroresonant element of n inputs, which jumps into the 'high' state by applying simultaneous pulses to the n inputs. The capacitance will be such as to give ferroresonance with only one of the inductors. On the other hand, this jump to the 'high' state will be impossible if the pulses, whatever their amplitude, apply to only n–1 inputs. This means that such a ferroresonant system behaves like an 'and' ferroresonant circuit of n inputs. The correct operation of such a circuit has been verified experimentally with up to five inputs, but we do not consider this number to be the limit.

Along the lines of the previous reasoning, if the assembly of n inductors in series is connected to a suitable capacitor we can obtain a circuit having a '1' signal in the output when there are '1' signals in at least m inputs, with $1 < m < n$ and m fixed by the value of the capacitor. In each case this capacitor is designed so that the ferroresonant curve obtained when polarizing m or more inductors is of the type given by curve (d'') (Fig. 5) and of the type given by curve (d') when only m–1 or less are polarized.

In this way some of the symmetrical functions can be carried out directly. The case $m = n$ would represent the n input 'and' circuit, which we have just described, and the case $m = 1$ would be that of an 'or' circuit. However, as we shall show in the following Section, the latter can also be obtained in an alternative method having several advantages.

(3.2) 'Or' Ferroresonant Circuit

It has been verified experimentally that the voltage/current characteristic of a set of n parallel equal non-linear inductors can be obtained from its individual characteristics by adding the currents corresponding to each voltage of the individual charac-

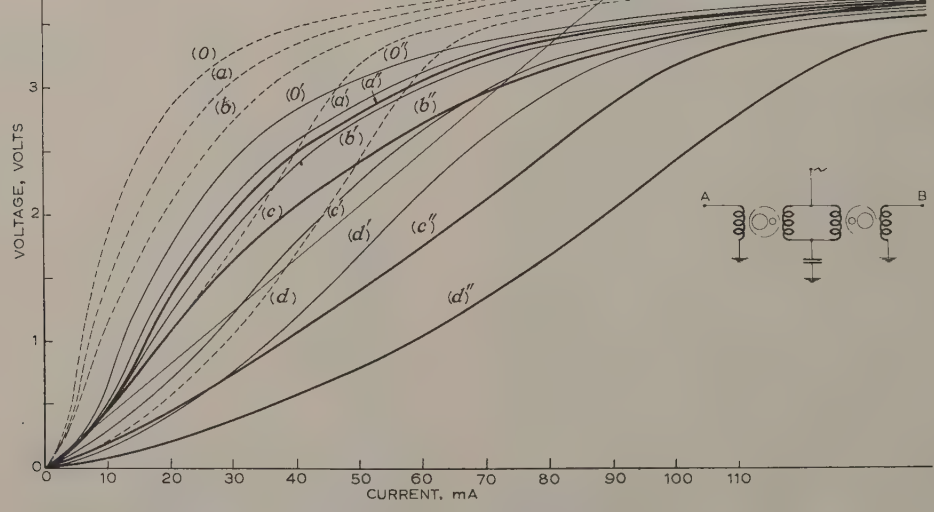


Fig. 6.—Voltage/current characteristics of a system of two equal inductors coupled in parallel.

teristics, the sum being the current corresponding to that given for the characteristic of the set. Therefore, the knee of the characteristic will have the same voltage co-ordinate as, and n times the current co-ordinate of, the knee of the characteristic of only one of the inductors considered individually.

For instance, consider two parallel non-linear inductors, and again this reasoning is by no means restrictive. Fig. 6, curves (o)–(d), shows the characteristic of each inductor considered individually. On assembling two equal inductors in parallel, the characteristics of the set, with polarization in one of the windings but not in the other, are represented by curves (o')–(d'), and the characteristics with simultaneous polarization on both windings are represented by curves (o'')–(d'').

Therefore, if we assemble the ferroresonant set as illustrated in Fig. 6, where the capacitance is twice that necessary to give ferroresonance with only one of the inductors, the system will jump from the 'low' to the 'high' state if an appropriate pulse is sent through one winding or simultaneous pulses are sent through both control windings.

We could apply identical reasoning to a set of parallel equal inductors and reach the same conclusions. Therefore, a set of n equal parallel-connected inductors with a suitable capacitor will constitute a ferroresonant element of n inputs which jumps into the 'high' state by sending pulses through any one of the inputs. The capacitance will be n times larger than that which gives ferroresonance with just one of the inductors, i.e. such a system behaves as an 'or' ferroresonant circuit of n inputs. The correct operation of such a circuit has been verified experimentally with the same extension as outlined when describing the 'and' ferroresonant circuit.

(3.3) Association of 'Or' and 'And' Circuits

The association of the fundamental circuits described can be carried out by setting up a system which can transfer information from the outputs of some circuits to the inputs of others.

Generally this association is accomplished by arranging successive levels of 'or' and 'and' circuits. If for the moment we limit ourselves to the two-level-circuit case, we can arrange an 'or' association of 'and' circuits or vice versa. We shall select that method in which the simplified expression of the Boolean function, which we wish to carry out as a sum of products or a product of sums, is the minimum one. However, it is possible to achieve a ferroresonant element directly which can perform the function which requires these associations. To this end we shall employ a complex inductor obtained through the connection of inductors in series-parallel assemblies in such a way that the required input combination triggers the ferroresonant element.

In the simple two-input case, we have the two arrangements shown in Fig. 7. On the basis of the considerations we have advanced to explain the operation of the fundamental 'and' and 'or' circuits, we can arrive at the conclusion that the left-hand circuit will jump into the 'high' state when the pulses arrive simultaneously in the A and C inputs or else in the B and D. This behaviour is represented by showing that the circuit performs the $AC + BD$ function. In the same way the right-hand circuit will give the function $(A + B)(C + D)$.

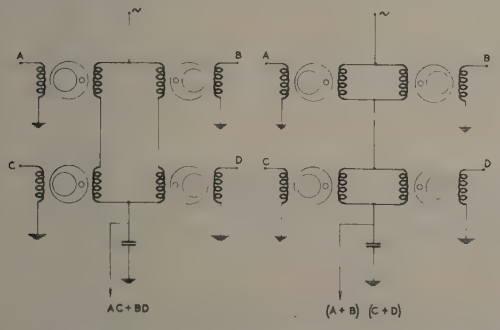


Fig. 7.—Ferroresonant two-level circuits.

However, the same number of inputs in each branch is not required. For instance, a circuit performing the function $A + BC$, including three inductors, will be discussed. Curves for the complex inductor have been drawn in Fig. 8 in agreement

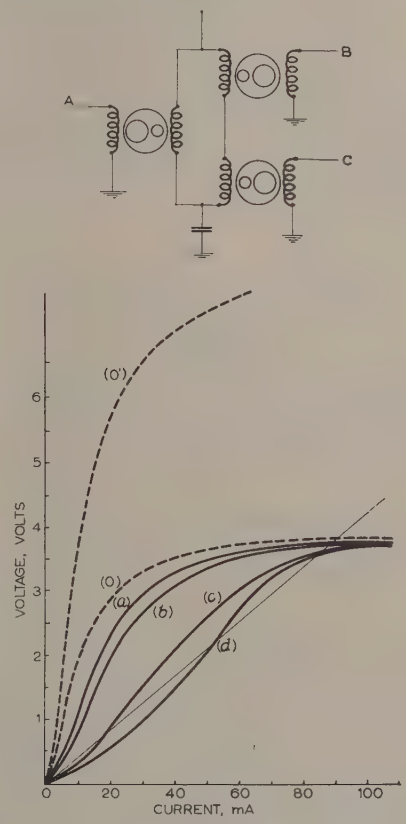


Fig. 8.—Voltage/current characteristic curves for the set of inductors coupled as shown.

with the previous considerations on the series-parallel association of inductors. Curves (o) and (o'), included for reference, have been taken from Fig. 4. The following curves have been obtained by adding currents for the A and the BC branches:

Curve (a), when no current circulates through any of the control windings.

Curve (b), when the control current reaches one and only one of the windings B or C.

Curve (c) corresponds to the case in which the polarizing current circulates simultaneously through windings B and C, but no current is sent through winding A.

Curve (d) corresponds to the case in which polarizing current is fed to input A but not to inputs B and C.

From these curves, and the capacitor characteristic included in Fig. 8, the ferroresonant curves illustrated in Fig. 9 have been plotted.

Let us assume that the generator voltage remains fixed at V; it is easily seen that the system will jump from the 'low' to the 'high' state only in the cases corresponding to curves (c) and (d),

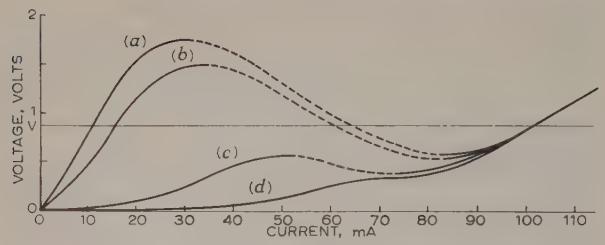


Fig. 9.—Ferroresonant curves for the inductors of Fig. 8 and a capacitor.

i.e. when polarization is effected through winding A only, or else through windings B and C simultaneously. Thus this circuit performs the function $A + BC$.

The satisfactory operation of this circuit and of that represented in Fig. 10 has been checked experimentally, but we do not consider that the number of inductors associated in them is a limiting figure.

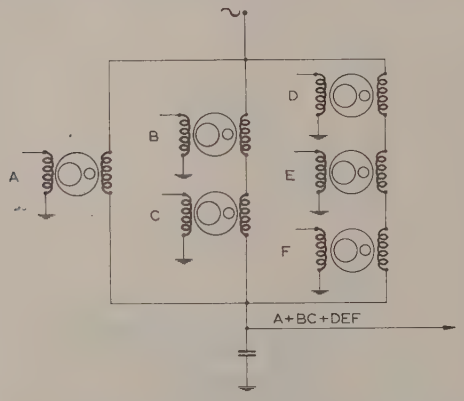


Fig. 10.—Association of three 'and' circuits in an 'or' circuit.

When the circuit structure demands that the same variable should reach the inputs of several inductors, the assembly could adopt a switching matrix arrangement; as it has been shown experimentally, a pulse in one of the input channels can trigger several cores if their control windings are arranged in series or parallel. When the method of shifting the information between elements described in connection with the shift register is used, it is also possible to branch the information in this manner.

Experimentally, we have proved that the transfer of information from one element to five others takes place easily when this method is used. This logical gain allows a rather flexible logical design. Furthermore, there is the possibility of using several independent shift pulse sequences properly timed with respect to the reset pulses. Logical gain will be in that case increased in the same relation.

The above allows for the development of a ferroresonant system of circuit logic setting up a complex inductor of a structure adequate to give a '1' output when the desired combination is present at the inputs. From the simplified Boolean function we can easily deduce the circuit structure. Of course, it would be necessary to assign input channels to variables and their complements, considering that the described ferroresonant circuits are not able to perform an inhibiting function.

(4) SYSTEM OF CIRCUIT LOGIC USING FERRORESONANT INHIBITION

It is also possible to carry out the 'inhibit' logical function through ferroresonant circuits. For this purpose it is necessary that the 'high' state of a ferroresonant element be capable of preventing the jump into the 'high' states of other elements even though the input combinations which would produce the jump appear in them.

A limiting coupling impedance, common to all elements, could perform this task if connected in series with the a.c. generator (Fig. 11). Every branch can be integrated by a complex ferro-

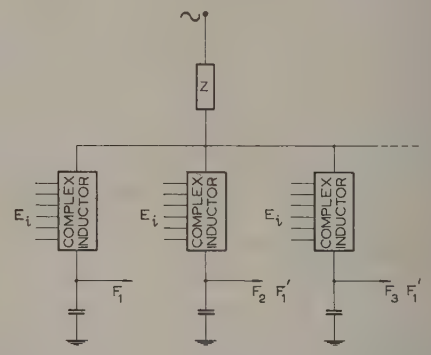


Fig. 11.—Inhibition between ferroresonant elements.

resonant element which carries out a specific function. The limiting impedance can be, for example, a capacitor. When the current which corresponds to the 'high' state of one of the ferroresonant elements flows through it, the drop in voltage must be enough to keep all the remaining branches from jumping into the 'high' state. In order that this coupling be possible it is necessary that the bistability regions of the different branches superimpose, allowing their connection to the same capacitor, and, furthermore, that the 'high'-state current of the inhibiting branch be equal to or higher than that in the remaining ones. These conditions can be obtained if we make the ferroresonant curves of all the elements which are to be coupled approximately the same. Furthermore, this results in the 'high' states being of the same order. We can do this by modifying the number of turns in the inductors as is shown in the case of the half-adder we shall describe later.

The use of this inhibition method allows us to modify the ferroresonant system of circuit logic described above. We can establish the condition for the inputs to the combinational circuits to be only the non-complemented variables. We can consider the possibility of writing a Boolean function as an expression satisfying the required condition, if we analyse this function written as a sum of products of the variables concerned. It can be seen immediately that each of the terms can be written as the product of two functions, one including all the non-complemented variables and the other the complemented ones. Morgan's theorem applied to the last function changes it into a function also including only non-complemented variables. Therefore, the complete Boolean function can be written in the form

$$F = \sum_{i,j} F_i(x, y, z, \ldots)F_j'(x, y, z, \ldots)$$

This sum will need a smaller or greater number of terms according to the simplification possibilities which the function offers. In some functions this sum can be reduced to just one term, and so

$$F = F_2(x, y, z, \ldots)F_1'(x, y, z, \ldots)$$

is a simple expression to be realized by a ferroresonant circuit of the type shown in Fig. 11. We must design the branch

corresponding to F_1 in such a way that, when it jumps into the 'high' state, it prevents the F_2 branch from jumping. Thus we obtain an inhibition, and the output taken from branch 2 will be the desired function $F_2F_1' = F$.

Furthermore, the circuit has another output, namely branch 1, in which function F_1 appears obviously with no inversion.

(5) APPLICATION TO SOME COMPUTING DEVICES

In this Section ferroresonant devices obtained from the circuits previously outlined are described. A half-adder and a series adder as examples of combinational circuits, and a shift register on which the adder is partly based, are also described.

(5.1) Half-Adder

Starting from the fundamental circuits we have outlined we can develop a half-adder. The outputs of this device can be written in Boolean algebra as

$$S = AB' + A'B$$
$$C = AB$$

To apply the inhibition system, we write the first expression as

$$S = (A + B)(AB)'$$

This function can be obtained by coupling a two-input 'or' circuit to a two-input 'and' circuit in such a way that the 'high' state in the 'and' circuit inhibits the 'high' state in the 'or' circuit (Fig. 12). The a.c. windings of the 'or' circuit have $2n$

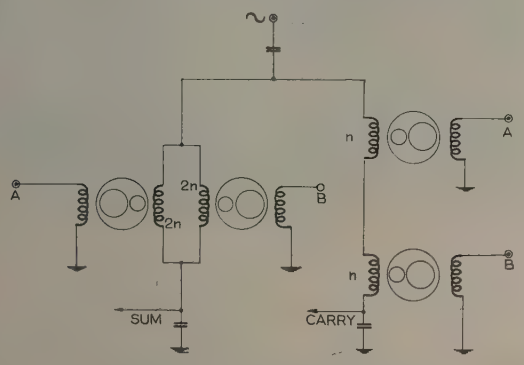


Fig. 12.—Ferroresonant half-adder.

turns and those of the 'and' circuit n turns, which give approximately equal ferroresonant curves in both ferroresonant branches, it being possible to supply them with the same alternating voltage. The capacitance, C, may easily be calculated to avoid both branches being in the 'high' state simultaneously. If we now sensitize the 'and' circuit more than the 'or' circuit, by changing C_{and} with respect to C_{or}, or by means of the input circuit to both control windings, or in both ways, there will be no doubt that the circuit outlined in Fig. 12 will behave as a ferroresonant half-adder.

The arrival of a '1' signal on channels A and B will make the 'and' circuit jump into the 'high' state while the 'or' circuit remains in the 'low' state. On the other hand, the arrival of a '1' signal on just one channel, and of a '0' signal on the other, will make the 'or' circuit jump into the 'high' state, while the 'and' circuit remains in the 'low' state. Finally, if '0' signals arrive through both channels, both branches will remain in the 'low' state. Therefore the sum digit is obtained from the capacitor corresponding to the 'or' circuit, and the carry digit from that corresponding to the 'and' circuit.

If, for any reason, the use of an equal number of turns in all a.c. windings is required, they may be coupled as outlined in Fig. 13 to ensure that both systems have an approximately equal ferroresonant curve. This method entails, however, the disadvantages of requiring twice as many inductors.

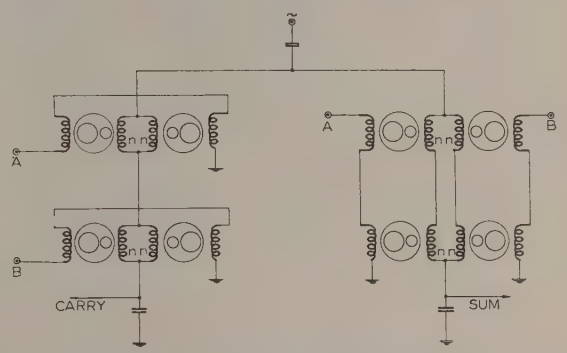


Fig. 13.—Ferroresonant half-adder with the same number of turns on cores.

(5.2) Shift Register

As was shown before, the basic ferroresonant element can behave as a delay flip-flop circuit if an intermediate storage circuit is added to it. We can obtain a shift register, the circuit of which is shown in Fig. 14, by joining several stages formed by this delay unit.

Each stage is connected to the intermediate data-storage circuit, establishing the link with the next each time a shifting pulse reaches C, which normally is polarized with a direct voltage, V_a. A is connected to a low output-impedance generator, which, by means of a pulse sequence σ_1, synchronized to the shifting pulses, σ_2, gives an amplitude- or frequency-modulated a.c. signal.

Suppose that a '0' signal is registered on branches F_1 and F_3, i.e. both are in the 'low' state, while on F_2 there is a '1'. Under these conditions, alternating voltage on P_1 and P_3 will be low and on P_2 high.

All ferroresonant elements will reset to the 'low' state when the drop in voltage reaches A so long as it drops below V_0 (Fig. 2). Then the direct voltage on Q_2, initially high, will decrease exponentially towards a low value, with a time-constant fixed at will.

When the shift pulse reaches C, the capacitor C_3' will be charged to a high value while the pulse lasts, and C_2' to a low value, at the expense of the acquired charges of C_1'' and C_2''. On the other hand, C_2' and C_3' will discharge with a damped oscillation, graduable at will, through the corresponding control windings. Since this discharge will produce a sufficient current to make the corresponding ferroresonant element jump to the 'high' state only when the acquired charge is high, in this case only branch 3 will jump to the 'high' state and so become the stored information shift.

The input of information in the first branch can also be controlled by the same shift sequence. It is, however, also possible to delay this input by a shorter time than a digit interval without disturbing the shift in the rest of the register.

The shift-register operation does not change if frequency modulation is used on the carrier instead of amplitude modulation.

If each of the stages of the shift register shown in Fig. 14 is substituted by a two-input 'or' circuit, as shown in Fig. 15, we

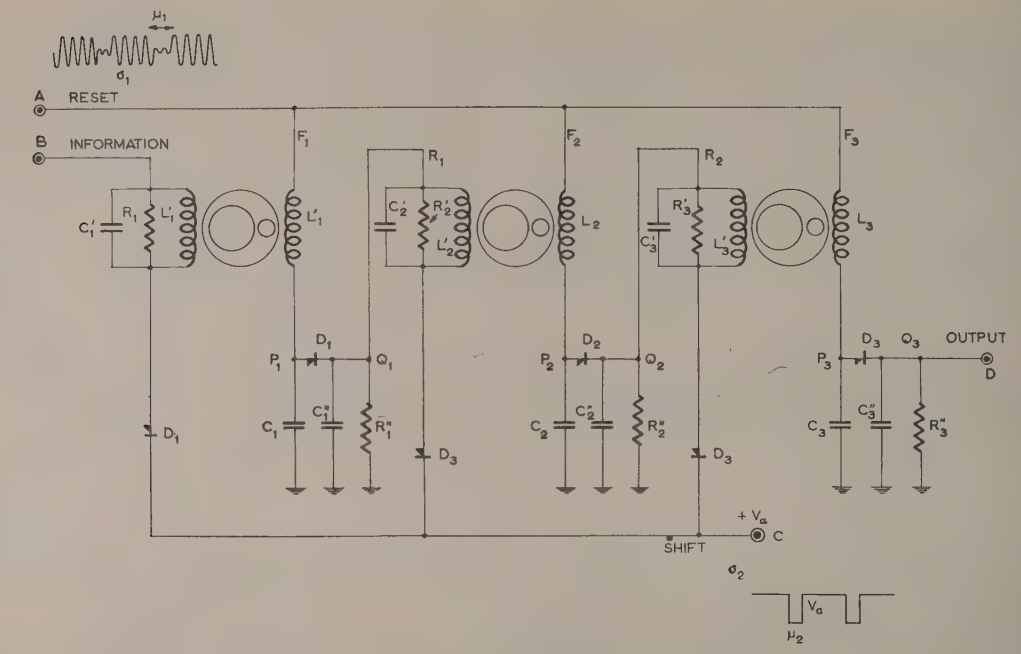


Fig. 14.—Shift-register circuit.

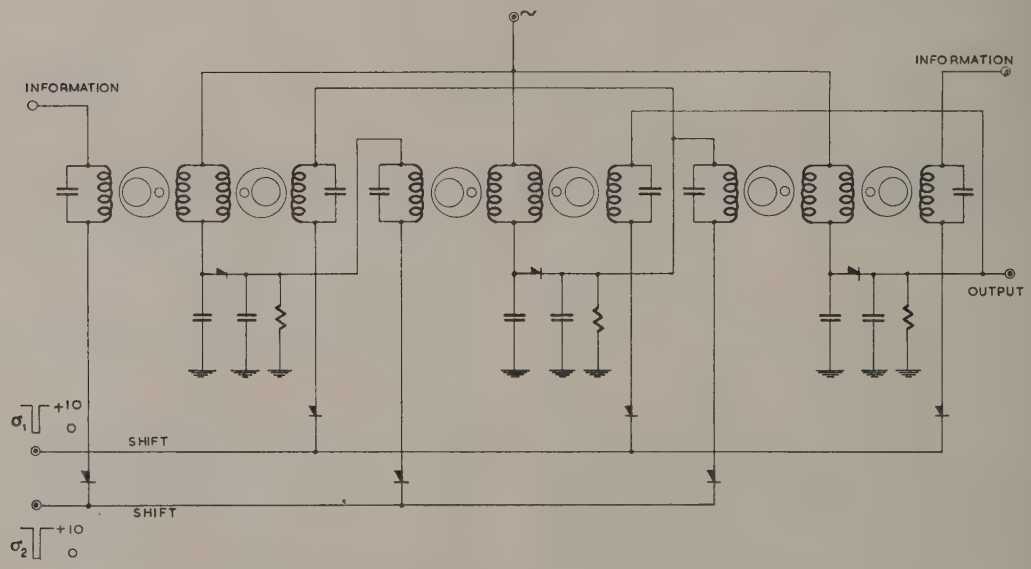


Fig. 15.—Shift register allowing the data shift in both directions.

obtain a 2-directional shift register. The σ_1 sequence will shift the stored information to the left and the σ_2 will shift it to the right.

(5.3) Series Full-Adder

It is also possible to develop a ferroresonant series adder from the half-adder and the aforementioned 'or' circuit. It will be enough to assemble two half-adders and one 'or' circuit, as shown in Fig. 16. For such an assembly we employ demodulating networks and a shift pulse sequence as described for the shift-register circuit. Such an 'or' circuit would have a double function: to obtain the sum of the two carry digits from the half-adders, and to produce the necessary delay for the result of this sum to reach the second half-adder during the next digit interval.

(5.4) Experimental Results

A typical ferroresonant element containing a Magnistor and a capacitor of 560 pF operates at a frequency of 2 Mc/s.

If the output of the generator is 4·5 volts (r.m.s.) a bistable

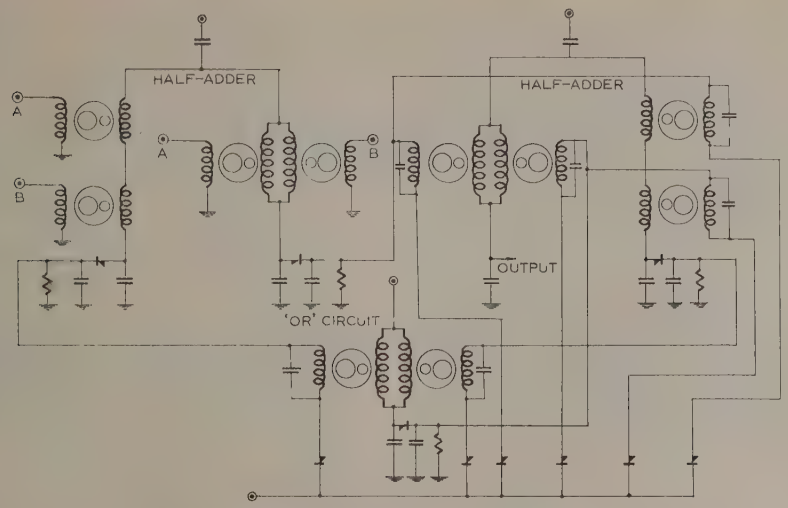


Fig. 16.—Ferroresonant series full-adder.

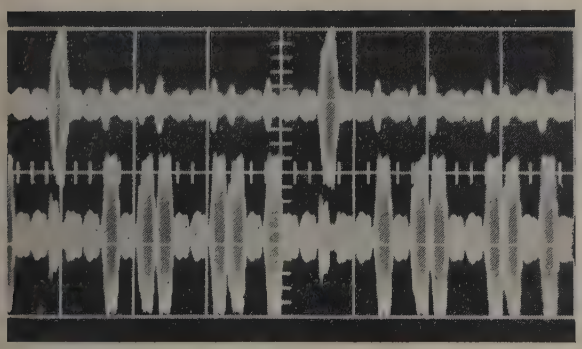


Fig. 17.—Oscillograms of behaviour of a half-adder operated with a carrier of 2 Mc/s.

(*a*) Carry.
(*b*) Sum.
Scales: 20 volts/cm;
20 microsec/cm.

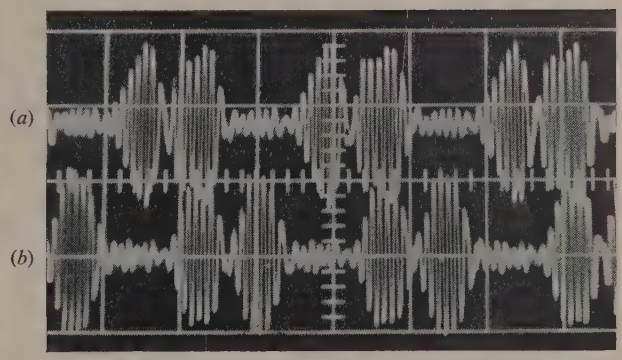


Fig. 18.—Oscillograms of shift register.

(*a*) 1st branch.
(*b*) 2nd branch.
Scales: 20 volts/cm;
5 microsec/cm.

region is obtained and the voltage across the capacitor is 15 volts in the 'high' state and 2·5 volts in the 'low' one. The generator voltage can be varied up to 25% without affecting the bistability.

The capacitor impedance leads to a current through this element of 100 mA in the 'high' and 17 mA in the 'low' states, with corresponding power requirements of 450 mVA and 77 mVA. This power is almost entirely reactive.

The jump transients are rather good: they last about two or three cycles of the carrier. The maximum repetition rate of the pulses is thus dependent on the carrier frequency.

The oscillograms shown in Fig. 17 correspond to the behaviour of a half-adder operated with a carrier of 2 Mc/s, and a reset pulse repetition rate of 2×10^5 pulses/sec. In this case the information entering channel A is 1000010000, and channel B, 100100.

The oscillograms shown in Fig. 18 correspond to the operation of a shift register. The carrier is 2 Mc/s and the reset and shift sequences have a repetition rate of 25×10^4 pulses/sec. The information shifted is 110110.

The time and voltage scales in the oscillograms allow us to confirm the previous statements made on the transient behaviour and the jump ratio of the elements.

(6) ACKNOWLEDGMENT

We wish to thank the United States Air Force, through the European Office Air Research and Development Command, for their support in the development of this research.

(7) REFERENCES

(1) BETHENOD, J.: 'Sur le transformateur à résonnance', *L'Éclairage Électrique*, 1907, **53**, p. 115.

(2) MARTIENSEN, O.: 'Uber neue Resonanzercheinungen in Wechselsstromkreisen', *Physikalische Zeitschrift*, 1910, p. 448.

(3) BOUCHEROT, P.: 'Résonnance électrique dans un circuit dont la self-induction contien du fer', *Comptes Rendus des Séances de l'Académie des Sciences*, 1920, **170**, p. 725.

(4) ROUELLE, R.: 'Contribution à l'étude expérimental de la ferro-résonnance', *Revue Générale de l'Electricité*, 1934, **36**, p. 715.

(5) SANTESMASES, J. G.: 'Sobre Ciertos Aspectos de la Ferro-rresonancia de un Circuito', *Anales de la Real Sociedad Española de Física y Química*, 1942, **38**, p. 21.

(6) ISBORN, C.: 'Ferroresonant Flip-Flops', *Electronics*, 1952, **25**, p. 121.
(7) SANTESMASES, J. G.: 'Fundamental Principles of Parallel Ferroresonance', Harvard University Computation Laboratory, 1952, Progress Report No. 22.
(8) SANTESMASES, J. G.: 'Circuito Disparador Basado en la Ferrorresonancia en Paralelo', *Anales de la Real Sociedad Española de Física y Química*, 1952, **48** A, p. 355.
(9) SANTESMASES, J. G.: 'Parallel Ferroresonant Triggers', Automatic Digital Computation, Proceedings of a Symposium held at the National Physical Laboratory, March, 1953.
(10) SANTESMASES, J. G., VIDAL, R. M., and SANCHEZ RODRIGUEZ, J.: 'Circuito Disparador Basado en la Ferrorresonancia en Paralelo, II—Uso de Ferritas', *Anales de la Real Sociedad Española de Física y Química*, 1954, **50** A, p. 47.
(11) DUINKER, S.: 'A Shifting Register using Ferroresonant Flip-Flops', *Applied Sciences Research*, 1955, **B–4**, p. 317.
(12) CUTLER, P. H.: 'Ferroresonant Trigger Circuits', S.R.D.E. Report No. 1083, 1955.
(13) SANTESMASES, J. G., and VIDAL, R. M.: 'Circuit déclencheur ferroresonnant parallele a rèaction', *L'Onde Electrique*, 1955, **35**, p. 165.
(14) ARBON, R. S., and JONES, P. H.: 'Ferroresonant computing circuits', *Wireless World*, 1956, **62**, p. 324.
(15) BELL, D. A., and NEWPORT, C. B.: 'Ferroresonant Circuits for Digital Computers', *Journal of the British Institution of Radio Engineers*, 1957, **17**, p. 619.
(16) SANTESMASES, J. G.: 'Switching Research in Spain', Symposium on the Theory of Switching, Harvard University, 1957.
(17) PROEBSTER, W. E.: 'Ferroresonant Switching Circuits', Congreso Internacional de Automática, Madrid, 1958.
(18) SANTESMASES, J. G.: 'Research on Ferroresonant Computer and Control Devices', U.S. Air Force Contract No. 61 (514) 1234, 1958.

21.387 : 621.316.722

The Institution of Electrical Engineers
Paper No. 3175 E
Mar. 1960
©

IMPEDANCE/FREQUENCY CHARACTERISTICS OF GLOW-DISCHARGE REFERENCE TUBES

By F. A. BENSON, D.Eng., Ph.D., Associate Member, and P. M. CHALMERS, B.Eng., Graduate.

(*The paper was first received 6th May, and in revised form 29th September,* 1959.)

SUMMARY

The impedance/frequency characteristics of some cold-cathode glow-discharge tubes over the frequency range 300 c/s–5 Mc/s have been obtained. The results are shown to agree with an extended form of the theory by van Geel. Impedance measurements indicate that there are several delayed effects in the glow discharge which affect the low-frequency impedance characteristics considerably. Tentative suggestions about the origin of these effects are discussed and the influence of some tube parameters on the magnitudes of the consequent delay times are indicated. The equivalent circuit of the glow-discharge tube is determined from the results.

LIST OF PRINCIPAL SYMBOLS

x = Distance measured from the cathode.
d = Anode-cathode spacing.
t = Time.
γ = Secondary-emission coefficient (Average number of electrons released at the cathode per incident positive ion).
α = First Townsend coefficient, i.e. the number of new electrons (or ions) produced by one electron in travelling a distance of 1 cm through the gas.
q = Regeneration factor defined as $\gamma\left[\exp\left(\int_0^d \alpha dx\right) - 1\right]$
($q = 1$ for a self-maintained discharge in the steady state with no superimposed alternating signal).
F = Transport factor = Q_d/Q_0. ($F = 1$ in a self-maintained discharge in the steady state.)
Q_d = Charge received at the anode when the cathode gives a charge Q_0 to the gas.
Q = Charge.
Q_1 = Total electron space charge.
I_d = Convection current at anode.
I_0 = Convection current at cathode.
I_1 = Electronic current.
I_2 = Positive-ion current.
I_{10} = Value of I_1 at the cathode.
I = Current.
V = Voltage.
G = Continuance factor = I_d/I_0.
ω = Angular frequency.
$D = d/dt$.
m_1, m_2, n_1, n_2 = Dimensionless constants.
$\phi_1 = (\partial F/\partial V)/(dF/\partial V)$
$\phi_2 = (\partial F/\partial I)/(dF/\partial I)$
$K_1 = (\partial Q/\partial V)/(dQ/\partial V)$
$K_2 = (\partial Q/\partial I)/(dQ/\partial I)$
} = Dimensionless constants.
(If the influence of the delayed-effect process is small ϕ_1, ϕ_2, K_1 and K_2 all approach unity.)
τ = Degression time = $-\frac{1}{I}\frac{dQ/\partial V}{dF/\partial V}$.
(This is associated with the time taken for the space charge to change.)
$R = -(dF/\partial I)/(dF/\partial V)$.
(This can be shown to be the slope of the static voltage/current characteristic.)
L = Coefficient of self-inductance for the glow discharge
$= \frac{1}{I}\frac{dQ/\partial I}{dF/\partial V}$.

(1) INTRODUCTION

When a small alternating voltage is superimposed on the steady operating voltage of a glow-discharge tube, the impedance presented to the alternating signal is found to increase with frequency. In certain glow-discharge tube applications, e.g. switching elements in automatic telephony, it is desirable to have a low audio-frequency impedance, and in some voltage-stabilizer circuits the transient response of the glow-discharge tube is an important factor which can be assessed if the equivalent circuit of the tube is known. The impedance has inductive and resistive components at audio frequencies, but the reactance is capacitive at frequencies of the order of 1 Mc/s. In order to understand the impedance/frequency characteristics better, in the present investigation the frequency range is extended to 5 Mc/s.

Previous measurements[1] to determine the a.f. impedance/frequency characteristics of tubes with various neon-argon fillings showed that the effective series inductance decreased and the resistance increased with increasing frequency over the range 300 c/s–30 kc/s, and both components decreased with increasing tube current.

Various attempts have been made to explain such results qualitatively in terms of the basic process in a glow discharge. It has been suggested[2] that the inductive component of the impedance is associated with the nature of conduction in the discharge, i.e. as the current varies the degree of ionization in the gas changes correspondingly, but there is a delay, due to the small mobility of the positive ions, before the degree of ionization reaches the new value. Williams,[3] however, states that the alternating electric field will have very little effect on the maintenance of the discharge and he attributes the inertia effect to the damped motion of current carriers. A charged particle moving in a steady electric field with a small alternating field superimposed produces in the external circuit a current which depends only on the average drift velocity of the particle. An approximate expression[3] for the drift velocity of a charged particle moving in a gas in terms of the collision frequency, particle mass and alternating field can be shown to be analogous to the expression for the current which flows in an electrical network containing inductance and resistance in series due to a given applied voltage. Williams's considerations, however, do not explain the impedance characteristics at higher frequencies and they necessitate the introduction of a capacitance, greater

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Benson and Mr. Chalmers are in the Department of Electrical Engineering, University of Sheffield.

than the interelectrode one, to account for the impedance peak observed at high frequencies.

An early theory by van Geel[4] was derived by considering perturbations about the steady-state maintenance condition and by assuming that physical quantities in the discharge (e.g. space-charge, etc.) can be regarded as functions of the instantaneous current and voltage alone. This theory predicted that the impedance loci should be semicircles if the displacement current is neglected.[5] Van Geel[6] has since extended the theory and postulated the form of the impedance locus when a secondary mechanism having a delay time is present in the discharge. The secondary mechanism, which is here termed a delayed effect, was called an after-effect by van Geel. The physical quantities in the discharge can now be regarded as functions of the instantaneous current and voltage and of a parameter, P, which is considered to be a retarded function of current and voltage. This parameter is discussed in Section 2. Van Geel's theory has not previously been confirmed experimentally, although Verhagen[5] has plotted impedance loci for tubes at frequencies up to 15 kc/s.

It has been suggested by Yeh[7] that there may be a correlation between the frequency at which the glow-discharge tube reactance changes over from inductive to capacitive and the positive-ion transit time in the cathode-fall region of the discharge. Benson and Gillespie[8] have shown how the transit time of the positive ions may be calculated if certain parameters of the discharge are known, but they had difficulty in comparing their experimental results with the theory because reliable figures for the various constants involved were not available for their particular tubes.

More recently, Ahsmann and Oskam[9] have shown that the impedance of helium discharge tubes at audio frequencies is reduced by additions of small quantities of neon, and a correlation between the self-inductance and the mobility of the dominant ions in the discharge has been suggested. The impedance loci they plotted are nearly semicircles with centres on the real axis, but no detailed analysis appears to have been made to show that the results are in full agreement with van Geel's theory.

In the present investigation it was decided to measure the impedance characteristics over the range 300 c/s–5 Mc/s in order to obtain the equivalent circuit of a glow-discharge tube and to test the validity of Yeh's suggested correlation. The tubes examined consisted of some with pure-neon fillings at various pressures, and others with various neon-argon mixtures. Each tube had a molybdenum rod anode 1 mm in diameter, and a cylindrical cathode 7·5 mm in diameter and 10 mm long. The electrode structure was mounted in a glass tube having a volume of about 8 cm^3. The tubes were sputtered until a layer of molybdenum, removed by bombardment from the cathode, covered the walls of the glass envelope adequately. The molybdenum layer shields the discharge from impurities normally released from the glass walls and also acts as a getter for certain contaminants, so that discharge tubes produced in this way generally give reliable and reproducible results.

(2) THEORETICAL TREATMENT

The impedance locus of a glow-discharge tube has the same form as that predicted by van Geel's theory[6] if it is extended to include several delayed effects. In the first treatment of the problem van Geel showed that, if the physical quantities of a discharge are regarded as functions of the instantaneous tube current and voltage, the impedance locus is a semicircle.[4, 6] In most glow discharges, however, there are secondary processes of electron production, e.g. the Penning effect,[10, 11] which can be regarded as retarded functions of current and voltage since they are associated with a delay time. Diffusion processes in the discharge can also be treated as delayed effects and White[12] has compared the time-constant of the semicircle impedance locus for a hollow-cathode tube with the ambipolar diffusion time of neon ions in the discharge. Such processes will affect the form of the tube-impedance locus considerably, as discussed in Section 5.2; hence, any process which affects the maintenance of the discharge and has a time-constant associated with it will be treated as a delayed effect.

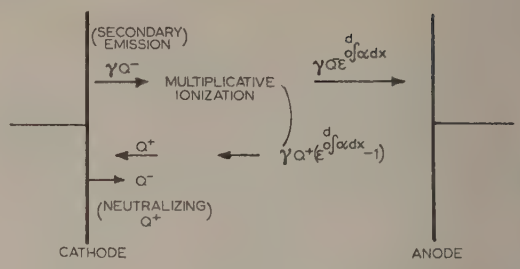


Fig. 1.—Charge transfer in a glow discharge.

The impedance of the discharge tube is derived in Section 9.1 by considering the effect of two delayed processes and eqn. (29) shows that the impedance has the following form:

$$Z = a_0 - \frac{a_1}{1 + j\omega\tau_1} - \frac{a_2}{1 + j\omega\tau_2} - \frac{a_3}{1 + j\omega\tau_3}$$

Impedance loci satisfying the above equation can be represented by the sum of three vectors whose individual loci are semicircles of diameters a_1, a_2 and a_3; for positive values of a_1, a_2, a_3 and τ_1, τ_2, τ_3 and $\tau_1 > \tau_2 > \tau_3$ the locus is shown in Fig. 2 and the construction is given in Section 9.3. If a_3 is negative the

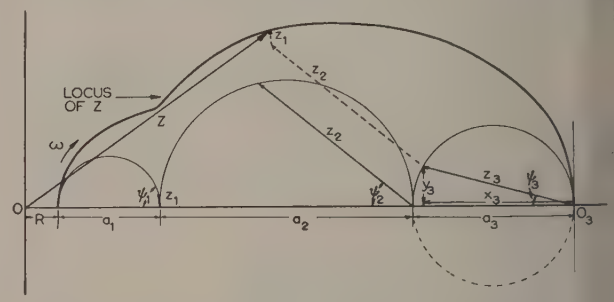


Fig. 2.—A theoretical impedance locus constructed for positive values of a_1, a_2 and a_3.
$\tau_1 : \tau_2 : \tau_3$ as 30 : 3 : 1

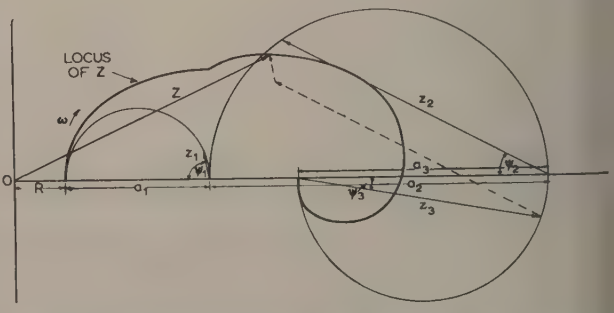


Fig. 3.—A theoretical impedance locus constructed for a negative value of a_3 and positive values of a_1 and a_2.
$\tau_1 : \tau_2 : \tau_3$ as 30 : 3 : 1

resultant locus and its components are as shown in Fig. 3. If the theory is extended to include n delayed effects, then

$$Z = a_0 - \left(\frac{a'}{1 + j\omega\tau'}\right) - \sum_{k=1}^{k=n}\left(\frac{a_k}{1 + j\omega\tau_k}\right)$$

The second term in this expression can be regarded as the contribution from the discharge process itself, and the last term as the contributions from the delayed-effect processes.

(3) MEASUREMENTS

The tube impedance over the range 300 c/s–30 kc/s was determined with the aid of the modified form of Owen bridge previously described.[1] The amplitude of the superimposed alternating current was held constant at 0·33 mA throughout the measurements, this value being chosen as a compromise to give an easily detectable null point with minimum distortion.

The impedance in the range 100 kc/s–5 Mc/s was measured on a r.f. bridge which indicated the equivalent parallel resistance and reactance and was supplied with an alternating voltage from a signal generator modulated at 1 kc/s. The null point was detected by a sensitive, well-shielded radio receiver. The bridge was a standard commercial one employing a multi-ratio tapped transformer principle. Fig. 4 shows the complete glow-discharge

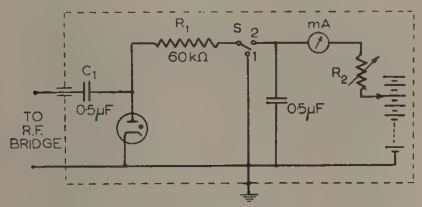


Fig. 4.—Glow-discharge-tube circuit for impedance measurements.

tube circuit, the components being mounted in a well-earthed metal box. The direct current through the tube was monitored and could be varied by changing the battery voltage and resistor R_2. The blocking capacitor, C_1, which was necessary to isolate the direct current from the bridge circuit, had negligible reactance compared with the tube impedance at 100 kc/s, and the series resistor, R_1, was non-inductive.

The parallel resistance and reactance of the tube were obtained by a difference method, thus eliminating stray capacitances and leakage resistances. The bridge was first balanced with the switch S in the earth position 1 and was then rebalanced with the tube switched on. These two readings enabled the equivalent parallel reactance and resistance to be evaluated. When the tube was replaced by known resistors and capacitors the maximum error of the measurements was found to be about 2%. Noise from the receiver and glow-discharge tube[13] masked the null point somewhat at frequencies below 500 kc/s but, even so, capacitance changes of 0·01 pF could be detected.

(4) RESULTS

Impedance loci were plotted from the results of the bridge measurements to facilitate the analysis of the results and to compare them with the theoretical predictions. If X_p and R_p are the measured values of parallel reactance and resistance of the tube, the corresponding series elements, X_s and R_s, are given by

$$X_s = X_p/[1 + (X_p/R_p)^2]$$

and

$$R_s = R_p/[1 + (R_p/X_p)^2]$$

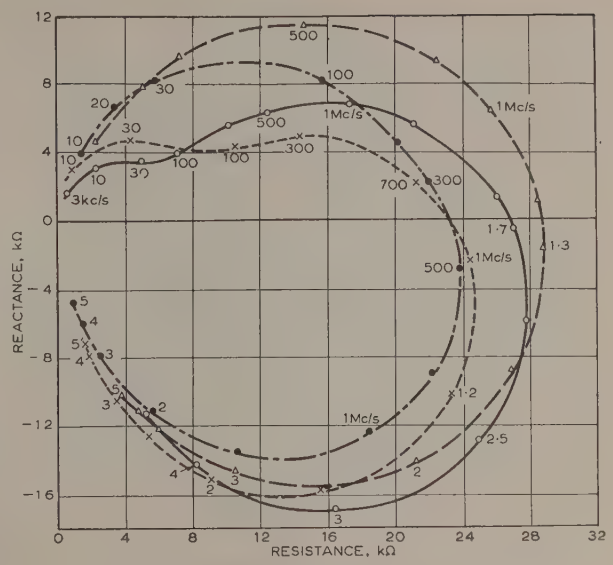


Fig. 5.—Typical impedance loci for tubes containing neon-argon mixtures.

—○— 10⁻³% argon.
--×-- 10⁻²% argon.
--●-- 0·1% argon.
—Δ— 1·0% argon.

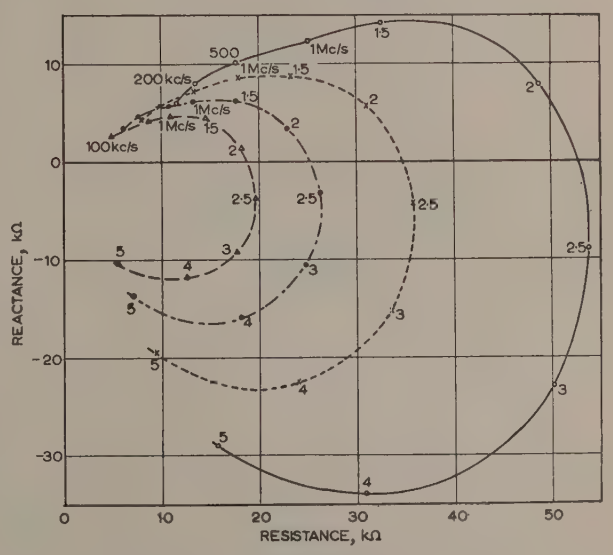


Fig. 6.—Impedance loci for a pure-neon tube operated at several values of tube current.

—○— 2 mA.
--×-- 3 mA.
--●-- 4 mA.
—Δ— 5 mA.

Fig 5 shows typical impedance loci for tubes having various argon contents for a tube current of 3 mA. Only one tube characteristic has been shown, since the spread in the characteristics of the tubes having a particular gas filling was small and the form of the impedance locus for frequencies above 10 kc/s was unaltered. Fig. 6 depicts the effect of varying the tube current on the form of the impedance locus for a pure-neon filling. Some of the 'pure'-neon tubes first examined were thought to be contaminated with a small trace of mercury vapour (0·002%)

although they had the same running and striking voltages as pure-neon tubes.[14] This impurity could affect the high-frequency characteristics of the tubes, so the present tubes were carefully prepared, and the possible contamination with mercury vapour was eliminated by using a suitable cold trap on the pumping system. The effect of the mercury vapour, however, was found to be very small. Some additional tubes having anode-cathode spacings varying between 2 and 4·5 mm were also examined at this stage and the impedance was found to be independent of the spacing.[15] Thus the characteristics of the tubes shown are for a normal glow-discharge which is not obstructed and does not have an anode-fall region.[16]

The effective parallel capacitance of all the tubes approached a limiting value at frequencies above 4 Mc/s (higher in the case of pure-neon tubes of 40 mm pressure and pure-argon tubes), seemingly independent of frequency. This capacitance, which was found to be independent of the anode-cathode spacing and electrode geometry, was taken to be the effective capacitance, C_d, of the glow-discharge. It was found to be of the order of the capacitance of the cathode-dark-space region of the discharge.[17]

Fig. 7 shows the variation of the effective capacitance with tube current and gas pressure for pure-neon tubes, and Fig. 8 is a similar graph for tubes having neon-argon mixtures.

In order to compare the results with the theory, which does not take into account the displacement current, the effective tube capacitance has been subtracted from the measured values of capacitance, and the resulting impedance loci are replotted in Figs. 9–11. For clarity the low-frequency parts of the loci are shown separately in Figs. 12 and 13.

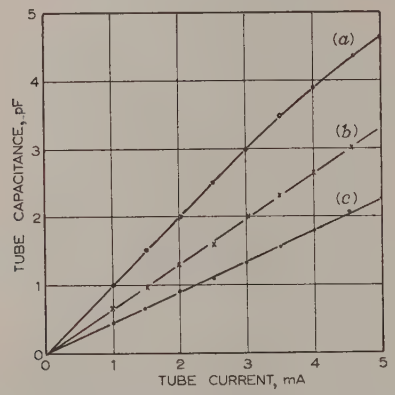


Fig. 7.—Variation of high-frequency capacitance with tube current and gas pressure for pure-neon tubes.

(*a*) 20 mm Hg.
(*b*) 30 mm Hg.
(*c*) 40 mm Hg.

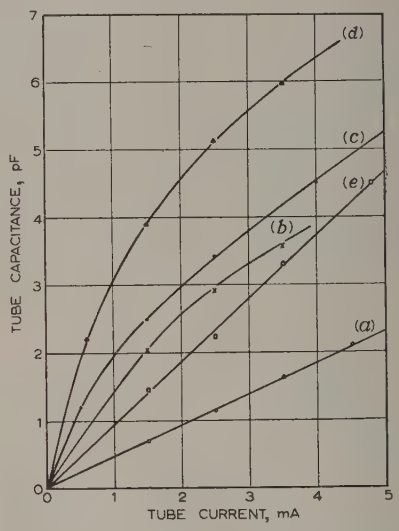


Fig. 8.—Variation of high-frequency capacitance with tube current for tubes containing neon-argon mixtures at 40 mm Hg total pressure.

(*a*) Pure neon.
(*b*) 10^{-3}% argon.
(*c*) 10^{-2}% argon.
(*d*) 0·1% argon.
(*e*) 1·0% argon.

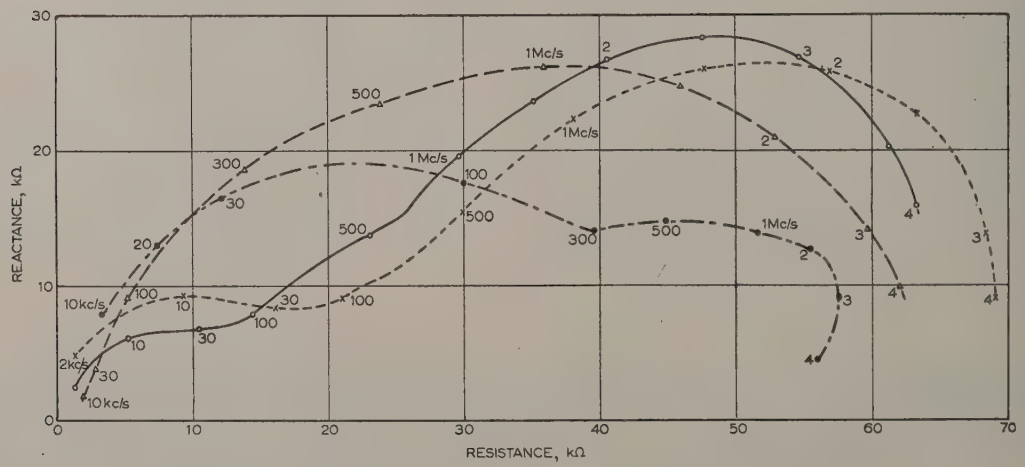


Fig. 9.—Modified impedance loci for tubes containing neon-argon mixtures.

—○— 10^{-3}% argon.
-- × -- 10^{-2}% argon.
—●— 0·1% argon.
—△— 1·0% argon.

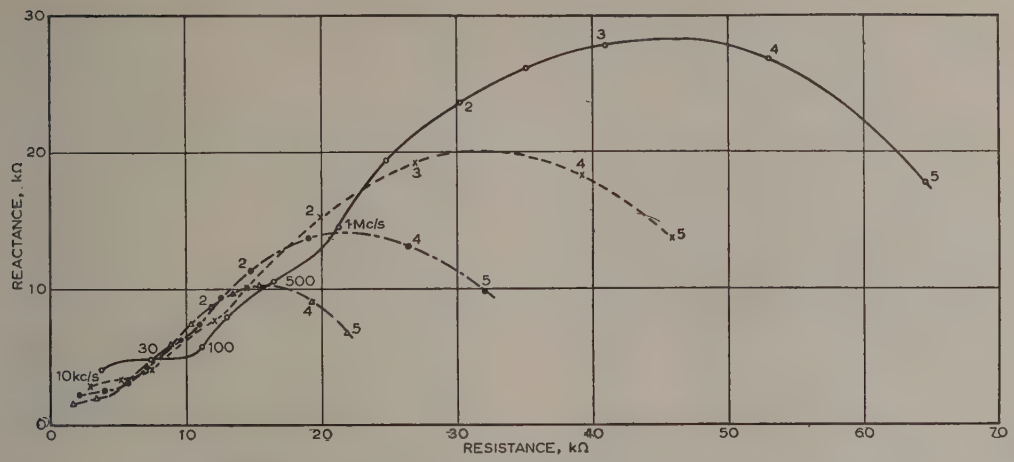


Fig. 10.—Modified impedance loci for a pure-neon tube operated at several tube currents.

—○— 2 mA.
--×-- 3 mA.
–-●–- 4 mA.
—Δ— 5 mA.

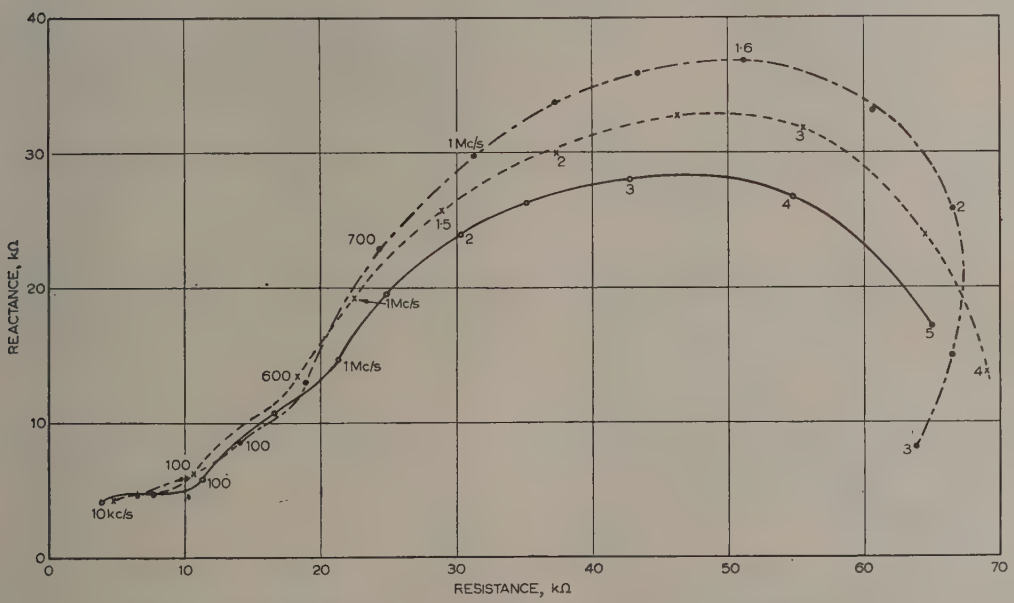


Fig. 11.—Modified impedance loci for pure-neon tubes having various gas pressures.

—○— 40 mm Hg.
--×-- 30 mm Hg.
–-●–- 20 mm Hg.

(5) DISCUSSION OF RESULTS

(5.1) High-Frequency Capacitance

The effective high-frequency capacitance of the pure-neon tube is proportional to the area of the cathode glow if the normal current density is assumed to be constant, and its value is less than the cathode-dark-space capacitance as calculated in Section 9.2. Van der Ziel and Chenette[17] have measured the admittance of some cold-cathode voltage-regulator tubes, and find the capacitance to be independent of frequency above 2 Mc/s. They also conclude that the capacitance is of the order of the dark-space capacitance. The positive-ion space-charge distribution and the electron space-charge in the negative-glow region could affect the capacitance considerably, however, and explain the difference between the measured value and the one obtained from the simplified calculation. Eqn. (31) nevertheless agrees in form with the observed variations of measured capacitance with current and pressure in the normal operating region.

With neon tubes containing small quantities of argon the current density is smaller and the normal-glow region is difficult to define.[18] The current density increases with increasing tube current, owing to cathode edge effects,[19] so that the capacitance will no longer be directly proportional to the current. Finally, in the abnormal-glow region, the current density will be approximately proportional to the current, and the dark-space distance is reduced at a rate less than the reciprocal of the current, so

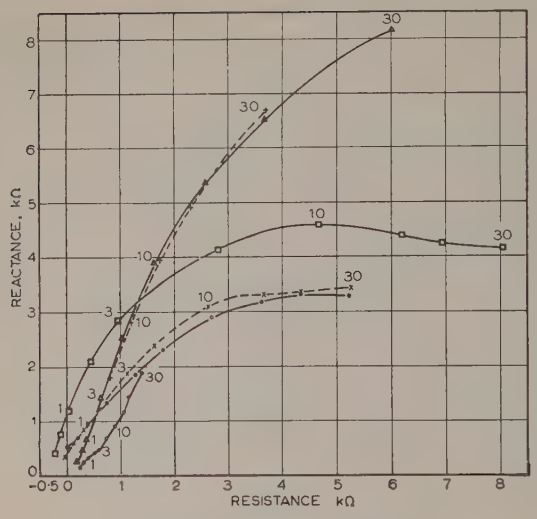


Fig. 12.—Audio-frequency portions of the impedance loci for tubes containing neon-argon mixtures.

—●— Pure neon.
--×-- 10^{-3}% argon.
—□— 10^{-2}% argon.
—△— 0·1% argon.
--+-- 0·3% argon.
—○— 1·0% argon.

Frequencies are in kilocycles per second.

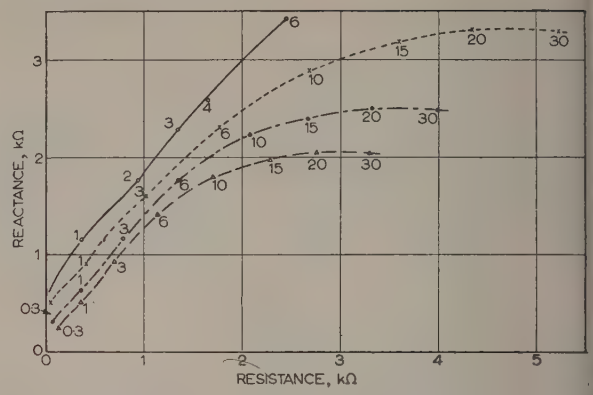


Fig. 13.—Effect of the tube current on the form of the audio-frequency impedance loci for a pure-neon tube.

—○— 2 mA.
--×-- 3 mA.
—●— 4 mA.
--△-- 5 mA.

Frequencies are in kilocycles per second.

(5.2) Impedance Loci

The impedance results for tubes having different anode cathode spacings seem to confirm the assumption that the plasma impedance can be neglected if there is no anode fall,[1] and therefore the impedance locus for the convection curren

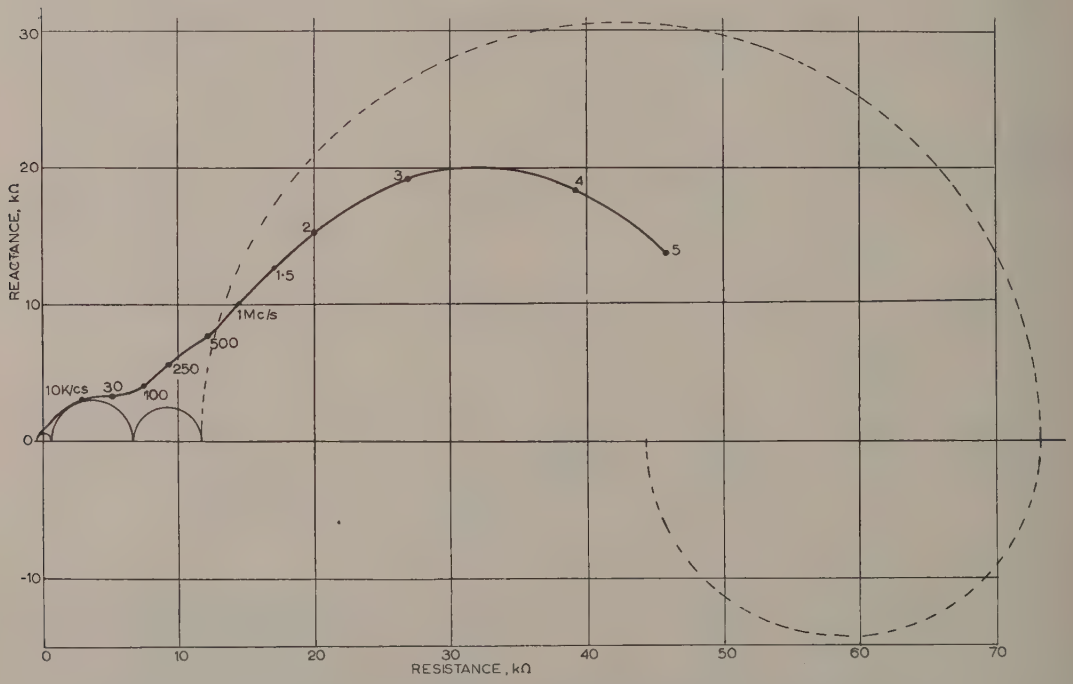


Fig. 14.—Modified impedance locus of a pure-neon tube showing component semi-circle loci.

$\tau_1 = 300\ \mu s$, $\tau_2 = 13\ \mu s$, $\tau_3 = 0{\cdot}64\ \mu s$,
$\tau_4 \simeq 0{\cdot}04\ \mu s$, $\tau_5 \simeq 0{\cdot}02\ \mu s$.
$a_1 = 1\ k\Omega$, $a_2 = 6\ k\Omega$, $a_3 = 5{\cdot}5\ k\Omega$,
$a_4 \simeq 62\ k\Omega$, $a_5 \simeq -30\ k\Omega$.

that the capacitance will change less with increasing tube current than in the normal-glow region. In tubes having 1% argon, the discharge again has a definable normal-glow region and the capacitance is directly proportional to the tube current.

can be derived by subtracting the high-frequency capacitance from the impedance.

The impedance loci of Figs. 9–11, when analysed according to the theory, can be represented by a series of vectors whose

ndividual loci are semicircles, each having a different time-onstant, τ. Fig. 14 shows how the measured impedance locus or a pure-neon tube can be represented. The loci for pure-eon tubes and neon-argon mixtures containing less than 0·1% rgon can be resolved into five such semicircular loci and there-ore into four active delay times. Tubes having more than ·1% argon have one less delay time, since in this case the delay me associated with the lifetime of the neon metastable atoms reduced to a negligibly small value by ionizing collisions with rgon atoms.

When analysing the impedance loci of all the tubes at high requencies in the above manner it is necessary to make the ector corresponding to the semicircle having the smallest time-onstant have a negative sign. The delay time associated with his locus is thought to be related to the positive-ion time. Thus in Figs. 5 and 6 the frequency at which the reactance is ero[7] will not necessarily be correlated with the ion transit time, ut will depend to some extent on the larger delay times which nfluence the form of the impedance locus at lower frequencies. The negative-resistance contribution due to the component ector locus, a_5, at frequencies greater than the reciprocal of on transit time is similar to the negative resistance exhibited by thermionic diode at very high frequencies.[20] The total ffective resistance will still be positive, however, since the lternating current in the discharge at these frequencies will be nainly due to electron motion and the electron transit time is everal orders smaller than the corresponding ion transit time. Further extension of the impedance locus at higher frequencies would be necessary to enable the value of the time-constant of he component locus, a_5, to be derived accurately.

In some cases it is difficult to determine the various component emicircular loci and time-constants of the measured impedance ocus by direct calculation or graphical construction, especially when the time-constants of consecutive loci have comparable values. It is proposed in some future work, however, to use a rial-and-error process involving the use of an analogue com-puter to simulate the glow-discharge tube equivalent circuit shown in Fig. 15) and the components L and a will be varied until the measured locus is approximated.

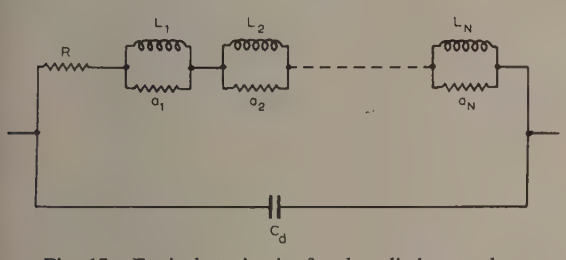


Fig. 15.—Equivalent circuit of a glow-discharge tube.

The brief analysis made on the results for a pure-neon tube indicate that the larger time-constants of the component semi-circular loci are independent of the tube current. The diameters of the loci are approximately inversely proportional to the tube current. This agrees with the theory if eqn. (28) is simplified by assuming $T \gg \tau$.

The larger delay times which affect the low-frequency impedance of the tubes could be due to the thermal lag associated with the heat-conduction mechanism from the cathode. The temperature variations in the cathode region will produce small changes in the gas density and consequently changes in the running voltage.[21]

Another possible delay time is that associated with the lateral growth of a glow-discharge.[22] The velocity with which the perimeter of the cathode glow expands if the current is suddenly increased is a function of the maintaining voltage of the glow discharge and approximate calculations indicate that the time associated with this process is of the order of 100 microsec.

(6) GENERAL CONCLUSIONS

The complete impedance locus of a glow-discharge tube can be explained effectively by van Geel's theory if extended to include several delay times. At frequencies above 3 Mc/s the effective tube capacitance is independent of frequency and is less than the cathode-dark-space capacitance (see Section 9.2). The delayed effects in a glow discharge affect the lower-frequency impedance considerably and the complete equivalent circuit is much more complicated than the one usually quoted.[23]

(7) ACKNOWLEDGMENTS

The work described has been carried out in the Department of Electrical Engineering at the University of Sheffield. The authors wish to thank Professor A. L. Cullen for facilities afforded in the laboratories of this Department and for his constant encouragement and advice. They also wish to acknowledge the kindness of the English Electric Valve Co. Ltd., in supplying all the special tubes for examination, for financial assistance with the work, for the loan of equipment, and for many helpful discussions and valuable suggestions.

(8) REFERENCES

(1) Benson, F. A., and Chalmers, P. M.: 'Effects of Argon Content on the Characteristics of Neon-Argon Glow-Discharge Reference Tubes', *Proceedings I.E.E.*, Monograph No. 321 R, December, 1958 (**106** C, p. 82).

(2) 'The Impedance of Voltage-Stabilizer and Reference Tubes', *Mullard Technical Communications*, 1954, **1**, p. 219.

(3) Williams, M. O.: 'Inertia Effects in Cold-Cathode Tubes', *Strowger Journal*, 1952, **8**, p. 106.

(4) van Geel, C.: 'Untersuchungen von Gasentladungen mit Rüchsicht auf ihre dynamischen Eigenshaften und ihre Stabilität', *Physica*, 1939, **6**, p. 806.

(5) Verhagen, C. J. D. M.: 'Impedanzmessungen an Gasentladungsröhen', *ibid.*, 1941, **8**, p. 361.

(6) van Geel, C.: 'Zelfinductie en Nawerking in Gasontladingen', Thesis submitted to the University of Delft, 1955.

(7) Yeh, C.: 'Note on Positive-Ion Transit Time in Glow-Discharge Tubes', *Journal of Applied Physics*, 1956, **27**, p. 98.

(8) Benson, F. A., and Gillespie, E. F. F.: 'Influence of Argon Content on the Characteristics of Glow-Discharge Tubes', *Proceedings I.E.E.*, Paper No. 2394 R, September, 1957 (**104** B, p. 498).

(9) Ahsmann, G., and Oskam, H. J.: 'Impedance and Recovery Time of Glow Discharges in Mixtures of Rare Gases', *Journal of Applied Physics*, 1958, **29**, p. 1768.

(10) Penning, F. M.: 'Über die Zündspannung von Ar–NO Gemischen', *Zeitschrift für Physik*, 1931, **72**, p. 338.

(11) Hoffman, A.: 'Die Bestimmung der Lebensdauer metastabiler Anregungszustände des Neon aus Reststrommessungen in Glimmentladungen', *ibid.*, 1942, **119**, p. 223.

(12) White, A. D.: 'New Hollow Cathode Glow Discharge', *Journal of Applied Physics*, 1959, **30**, p. 711.

(13) Benson, F. A.: 'Gas-Filled Voltage Stabilizers', *Electronic and Radio Engineer*, 1957, **34**, p. 16.

(14) Jurriaanse, T., Penning, F. M., and Moubis, J. H. A.: 'The Normal Cathode Fall for Molybdenum and Zirconium in the Rare Gases', *Philips Research Reports*, 1946, **1**, p. 225.

(15) AHSMANN, G.: 'On the Impedance of the Anode Fall Region of a Glow Discharge in the Noble Gases', Proceedings of the Third International Conference on Ionization Phenomena in Gases, Venice, 1957.
(16) TOWNSEND, M. A., and DEPP, W. A.: 'Cold-Cathode Tubes for Transmission of Audio-Frequency Signals', *Bell System Technical Journal*, 1953, **32**, p. 1371.
(17) VAN DER ZIEL, A., and CHENETTE, E. R.: 'Noise and Impedance Measurements in Voltage Regulator Tubes', *Physica*, 1957, **23**, p. 943.
(18) MEISSNER, J.: 'Über den Einfluss der metastabilen Anregungszustände auf die normale Stromdichte und den normalen Kathodenfall der Glimmentladung in Edelgasen und Edelgasgemischen', *Zeitschrift für Physik*, 1941, **117**, p. 325.
(19) SEELIGER, R., and REGER, M.: 'Über die Stromdichte des normalen Kathodenfalles', *Annalen der Physik*, 1927, **83**, p. 535.
(20) SPANGENBERG, K. R.: 'Vacuum Tubes' (McGraw-Hill, 1948), p. 486.
(21) JURRIAANSE, T.: 'The Influence of Gas Density and Temperature on the Normal Cathode Fall of a Gas Discharge in Rare Gases', *Philips Research Reports*, 1946, **1**, p. 407.
(22) VON ENGEL, A.: 'Ionised Gases' (Clarendon Press, 1955), p. 188.
(23) BENSON, F. A., and BENTAL, L. J.: 'Glow-Discharge Stabilizers', *Wireless Engineer*, 1955, **32**, p. 330, and 'Glow-Discharge Tubes', *ibid.*, 1956, **33**, p. 33.
(24) LOEB, L. B.: 'Fundamental Processes of Electrical Discharge in Gases' (John Wiley, 1939), p. 377.

(9) APPENDICES

(9.1) The Impedance of a Glow-Discharge Tube Exhibiting Two Delayed Effects

The following is an extension of van Geel's theory but is somewhat simplified by neglecting ionization due to positive ions (an assumption which is justifiable for the present glow-discharge tubes[24]).

Consider the electron multiplication process when one electron is released from the cathode of the tube. The number of new electrons formed by ionizing collisions in the distance d is given by $\exp \int_0^d \alpha dx$. The number of positive ions formed is $\exp\left(\int_0^d \alpha dx\right) - 1$. These ions travel to the cathode and release $\gamma\left[\exp\left(\int_0^d \alpha dx\right) - 1\right]$ electrons.

Assume that a positive charge Q is collected at the cathode. This is equivalent to the gas gaining a negative charge Q (see Fig. 1). The positive charge Q also gives rise to a negative charge γQ which is transferred to the gas. The total charge received by the gas is therefore $(1 + \gamma)Q$. The charge received by the anode is $\gamma Q \exp \int_0^d \alpha dx$. Hence,

$$F = \gamma \exp \int_0^d \alpha dx/(1 + \gamma) = (q + \gamma)/(1 + \gamma) \quad . \quad (1)$$

The relationship between G and F has been determined by van Geel[6] and it can be shown that

$$G = F - \frac{1}{I_0}\frac{dQ_1'}{dt} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where Q_1' is a function of the total electron space-charge, Q_1. If Q_T is the total resulting space-charge,

$$\frac{dQ_T}{dt} = I_d - I_0 = I_0(G - 1) \quad . \quad . \quad . \quad . \quad (3)$$

From eqns. (2) and (3),

$$I_0(F - 1) = \frac{dQ_T}{dt} + \frac{dQ_1'}{dt} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Writing $Q = Q_T + Q_1'$,

$$I_0(F - 1) = \frac{dQ}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Suppose the transport factor is dependent on two delayed effects P_1 and P_2, which have time-constants T_1 and T_2. P_1 and P_2 are considered as retarded functions of V and I. If I changes to $I + i$ and V to $V + v$, the corresponding changes in P_1 and P_2 may be written as $P_1 + P_1'$ and $P_2 + P_2'$ respectively.

When the new equilibrium condition of the discharge is reached, $P_1' = P_0$ and

$$P_0 = \frac{\partial P_1}{\partial V}v + \frac{\partial P_1}{\partial I}i \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Suppose that the current has changed before P' has reached the new equilibrium value P_0; then, in general,

$$P_1' + T_1\frac{dP_1'}{dt} = P_0 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

or

$$P_1' = P_0/(1 + T_1 D) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

From eqns. (6) and (8),

$$P_1' = \frac{\partial P_1}{\partial V}\left(\frac{v}{1 + T_1 D}\right) + \frac{\partial P_1}{\partial I}\left(\frac{i}{1 + T_1 D}\right) \quad . \quad . \quad (9)$$

Similarly,

$$P_2' = \frac{\partial P_2}{\partial V}\left(\frac{v}{1 + T_2 D}\right) + \frac{\partial P_2}{\partial I}\left(\frac{i}{1 + T_2 D}\right) \quad . \quad . \quad (10)$$

The change in F is given by

$$F - 1 = \frac{\partial F}{\partial V}v + \frac{\partial F}{\partial I}i + \frac{\partial F}{\partial P_1}P_1' + \frac{\partial F}{\partial P_2}P_2' \quad . \quad . \quad (11)$$

Substituting for P_1' and P_2' from eqns. (9) and (10),

$$F - 1 = \frac{\partial F}{\partial V}v + \frac{\partial F}{\partial I}i + \frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial V}\left(\frac{v}{1 + T_1 D}\right) + \frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial I}\left(\frac{i}{1 + T_1 D}\right) + \frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial V}\left(\frac{v}{1 + T_2 D}\right) + \frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial V}\left(\frac{i}{1 + T_2 D}\right) \quad . \quad (12)$$

Now

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial V}Dv + \frac{\partial Q}{\partial I}Di + \frac{\partial Q}{\partial P_1}DP_1' + \frac{\partial Q}{\partial P_2}DP_2' \quad . \quad (13)$$

Substituting again from eqns. (9) and (10),

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial V}Dv + \frac{\partial Q}{\partial I}Di + \frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial V}\left(\frac{Dv}{1 + T_1 D}\right) + \frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial I}\left(\frac{Di}{1 + T_1 D}\right) + \frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial V}\left(\frac{Dv}{1 + T_2 D}\right) + \frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial I}\left(\frac{Di}{1 + T_2 D}\right) \quad . \quad (14)$$

Also, the total change in F due to change in V gives

$$\frac{dF}{\partial V} = \frac{\partial F}{\partial V} + \frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial V} + \frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial V} \quad . \quad . \quad . \quad (15)$$

Similarly,

$$\frac{dF}{\partial I} = \frac{\partial F}{\partial I} + \frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial I} + \frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial I} \quad . \quad . \quad . \quad (16)$$

$$\frac{dQ}{\partial V} = \frac{\partial Q}{\partial V} + \frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial V} + \frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial V} \quad . \quad . \quad . \quad (17)$$

and

$$\frac{dQ}{\partial I} = \frac{\partial Q}{\partial I} + \frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial I} + \frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial I} \quad . \quad . \quad . \quad (18)$$

Let

$$\frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial V} = m_1\frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial V} \quad . \quad . \quad . \quad . \quad (19)$$

$$\frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial V} = m_2\frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial V} \quad . \quad . \quad . \quad . \quad (20)$$

$$\frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial I} = n_1\frac{\partial F}{\partial P_2}\frac{\partial P_2}{\partial I} \quad . \quad . \quad . \quad . \quad (21)$$

$$\frac{\partial Q}{\partial P_1}\frac{\partial P_1}{\partial I} = n_2\frac{\partial Q}{\partial P_2}\frac{\partial P_2}{\partial I} \quad . \quad . \quad . \quad . \quad (22)$$

Substituting in eqn. (12) from eqns. (15), (16), (19) and (21) gives

$$\begin{aligned} F - 1 = {} & \frac{\partial F}{\partial V}v + \left(\frac{dF}{\partial V} - \frac{\partial F}{\partial V}\right)\left(\frac{m_1}{1+m_1}\right)\left(\frac{v}{1+T_1D}\right) \\ & + \left(\frac{dF}{\partial V} - \frac{\partial F}{\partial V}\right)\left(\frac{1}{1+m_1}\right)\left(\frac{v}{1+T_2D}\right) + \frac{\partial F}{\partial I}i \\ & + \left(\frac{dF}{\partial I} - \frac{\partial F}{\partial I}\right)\left(\frac{n_1}{1+n_1}\right)\left(\frac{i}{1+T_1D}\right) \\ & + \left(\frac{dF}{\partial I} - \frac{\partial F}{\partial I}\right)\left(\frac{1}{1+n_1}\right)\left(\frac{i}{1+T_2D}\right) \quad . \quad . \quad . \quad (23) \end{aligned}$$

and substituting in eqn. (14) from eqns. (17), (18), (20) and (22),

$$\begin{aligned} DQ = {} & \frac{\partial Q}{\partial V}Dv + \left(\frac{dQ}{\partial V} - \frac{\partial Q}{\partial V}\right)\left(\frac{m_2}{1+m_2}\right)\left(\frac{Dv}{1+T_1D}\right) \\ & + \left(\frac{dQ}{\partial V} - \frac{\partial Q}{\partial V}\right)\left(\frac{1}{1+m_2}\right)\left(\frac{v}{1+T_2D}\right) \\ & + \frac{\partial Q}{\partial I}Di + \left(\frac{dQ}{\partial I} - \frac{\partial Q}{\partial I}\right)\left(\frac{n_2}{1+n_2}\right)\left(\frac{Di}{1+T_1D}\right) \\ & + \left(\frac{dQ}{\partial I} - \frac{\partial Q}{\partial I}\right)\left(\frac{1}{1+n_2}\right)\left(\frac{i}{1+T_2D}\right) \quad . \quad . \quad . \quad (24) \end{aligned}$$

Substituting for $F - 1$ and DQ in eqn. (5) and multiplying by $(1 + T_1D)(1 + T_2D)$,

$$\begin{aligned} & \frac{\partial F}{\partial V}\left[T_1T_2D^2v + \left(\frac{T_2 + m_1T_1}{1+m_1}\right)Dv\right] + \frac{dF}{\partial V}\frac{1}{(1+m_1)} \\ & \quad [m_1(1 + T_2D)v + (1 + T_1D)v] - \frac{1}{I}\left[\frac{\partial Q}{\partial V}T_1T_2D^3v \right. \\ & \quad \left. + \left(\frac{T_2 + m_2T_1}{1+m_2}\right)D^2v\right] + \frac{dQ}{\partial V}\left(\frac{1}{1+m_2}\right) \\ & \quad [m_2(D + T_2D^2)v + (D + T_1D^2)v] = \frac{1}{I}\left[\frac{\partial Q}{\partial I}T_1T_2D^3i \right. \\ & \quad \left. + \left(\frac{T_2 + n_2T_1}{1+n_2}\right)D^2i\right] + \frac{dQ}{\partial I}\left(\frac{1}{1+n_2}\right) \\ & \quad [n_2(D + T_2D^2)i + (D + T_1D^2)i] - \left[\frac{\partial F}{\partial I}T_1T_2D^2i \right. \\ & \quad \left. + \left(\frac{T_2 + n_1T_1}{1+n_1}\right)Di\right] + \frac{dF}{\partial I}\left(\frac{1}{1+n_1}\right) \\ & \quad [n_1(1 + T_2D)i + (1 + T_1D)i] \quad . \quad . \quad . \quad . \quad . \quad . \quad (25) \end{aligned}$$

Dividing by $dF/\partial V$ and rearranging,

$$\begin{aligned} & v + \left[\left(\frac{\partial F/\partial V}{dF/\partial V}\right)\left(\frac{T_2 + m_1T_1}{1+m_1}\right) + \left(\frac{m_1T_2 + T_1}{1+m_1}\right)\right]Dv \\ & \quad + \left(\frac{\partial F/\partial V}{dF/\partial V}\right)T_1T_2D^2v - \frac{1}{I}\left(\frac{\partial Q/\partial V}{dF/\partial V}\right)T_1T_2D^3v \\ & \quad - \frac{1}{I}\left[\left(\frac{\partial Q/\partial V}{dF/\partial V}\right)\left(\frac{T_2 + m_2T_1}{1+m_2}\right) + \left(\frac{\partial Q/\partial V}{dF/\partial V}\right)\right. \\ & \quad \left.\left(\frac{m_2T_2 + T_1}{1+m_2}\right)\right]D^2v - \frac{1}{I}\left(\frac{dQ/\partial V}{dF/\partial V}\right)Dv \\ & \quad = -\left(\frac{dF/\partial I}{dF/\partial V}\right)i - \left[\left(\frac{\partial F/\partial I}{dF/\partial V}\right)\left(\frac{T_2 + n_1T_1}{1+n_1}\right)\right. \\ & \quad \left. + \left(\frac{dF/\partial I}{dF/\partial V}\right)\left(\frac{n_1T_2 + T_1}{1+n_1}\right)\right]Di - \left(\frac{\partial F/\partial I}{dF/\partial V}\right)T_1T_2D^2i \\ & \quad + \frac{1}{I}\left(\frac{dQ/\partial V}{dF/\partial V}\right)Di + \left[\frac{1}{I}\left(\frac{\partial Q/\partial I}{dF/\partial V}\right)\left(\frac{T_2 + n_2T_1}{1+n_2}\right)\right. \\ & \quad \left. + \frac{1}{I}\left(\frac{dQ/\partial I}{dF/\partial V}\right)\left(\frac{n_2T_2 + T_1}{1+n_2}\right)\right]D^2i + \frac{1}{I}\left(\frac{\partial Q/\partial I}{dF/\partial V}\right)T_1T_2D^3i \\ & \quad \quad . \quad . \quad . \quad (26) \end{aligned}$$

Eqn. (26) can be written as

$$\begin{aligned} & v + \left[\phi_1\left(\frac{T_2 + m_1T_1}{1+m_1}\right) + \left(\frac{m_1T_2 + T_1}{1+m_1}\right)\right]Dv + \phi_1T_1T_2D^2v \\ & \quad + K_1\tau T_1T_2D^3v + \left[K_1\tau\left(\frac{T_2 + m_2T_1}{1+m_1}\right)\right. \\ & \quad \left. + \tau\left(\frac{m_2T_2 + T_1}{1+m_2}\right)\right]D^2v + \tau Dv \\ & \quad = Ri + \left[R\phi_2\left(\frac{T_2 + n_1T_1}{1+n_1}\right) + R\left(\frac{n_1T_2 + T_1}{1+n_1}\right)\right]Di \\ & \quad + \phi_2RT_1T_2D^2i + LDi + \left[K_2L\left(\frac{T_2 + n_2T_1}{1+n_2}\right)\right. \\ & \quad \left. + L\left(\frac{n_2T_2 + T_1}{1+n_2}\right)\right]D^2i + K_2LT_1T_2D^3i \quad . \quad . \quad . \quad (27) \end{aligned}$$

For sinusoidal variations, $D = j\omega$, hence the impedance $Z = v/i$ is

$$Z = \frac{R + \left\{\left[\phi_2\left(\frac{T_2 + n_1T_1}{1+n_1}\right) + \left(\frac{n_1T_2 + T_1}{1+n_1}\right)\right]R + L\right\}j\omega - \left\{\left[K_2\left(\frac{T_2 + n_2T_1}{1+n_2}\right) + \left(\frac{n_2T_2 + T_1}{1+n_2}\right)\right]L + \phi_2RT_1T_2\right\}\omega^2 - K_2LT_1T_2j\omega^3}{1 + \left[\phi_1\left(\frac{T_2 + m_1T_1}{1+m_1}\right) + \left(\frac{m_1T_2 + T_1}{1+m_1}\right) + \tau\right]j\omega - \left\{\left[K_1\left(\frac{T_2 + m_2T_1}{1+m_2}\right) + \left(\frac{m_2T_2 + T_1}{1+m_2}\right)\right]\tau + \phi_1T_1T_2\right\}\omega^3 - K_1\tau T_1T_2j\omega^3}$$

. . . (28)

Eqn. (28) can be reduced to

$$Z = a_0 - \frac{a_1}{1 + j\omega\tau_1} - \frac{a_2}{1 + j\omega\tau_2} - \frac{a_3}{1 + j\omega\tau_3} \quad . \quad . \quad (29)$$

where a_0, a_1, etc., have the dimensions of resistance; when the delayed effect is small and $\tau_1 \gg \tau_2 \gg \tau_3$, then τ_1 is associated with T_1, τ_2 is associated with T_2, and τ_3 is associated with τ.

(9.2) Effective Capacitance of the Glow-Discharge Tube

The capacitance of the glow discharge at high frequencies is

$$C_d \simeq \frac{\epsilon_0 a}{d_c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where a is the area of the discharge and d_c is the width of the cathode dark space.

If the similarity rules are valid for the discharge, pd_c and Jn/p^2 are constant and therefore eqn. (30) can be written

$$C_d \simeq KIp \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where K is a constant for a given gas filling and cathode material.

This was found to be true in the case of the pure-neon tubes operating in the normal-glow region, as shown in Fig. 7.

Taking $Jn = 5{\cdot}5\,\text{mA/cm}^2$, $d_c = 0{\cdot}023\,\text{cm}$,* $I = 3\,\text{mA}$ for a pure-neon filling at 40 mm pressure and a molybdenum cathode, C_d according to eqn. (1) is 2·09 pF. The measured value of the capacitance was 1·52 pF.

Space charges in the cathode-fall and negative-glow regions will modify the simplified calculations to some extent, and d_c should possibly be replaced by a distance l (greater than d_c) which will depend directly on d_c and the length of the negative glow, l_g. However, since pl_g is constant, the form of eqn. (31) will still be applicable.

(9.3) Construction of the Theoretical Impedance Loci and the Nature of the Equivalent Circuit of the Glow-Discharge Tube

The impedance locus of a glow discharge with two delayed effects will be taken as an example to illustrate the construction. From Section 9.1,

$$Z = a_0 - \frac{a_1}{1 + j\omega\tau_1} - \frac{a_2}{1 + j\omega\tau_2} - \frac{a_3}{1 + j\omega\tau_3} \quad . \quad (32)$$

or

$$Z = a_0 + z_1 + z_2 + z_3 \quad . \quad . \quad . \quad . \quad (33)$$

* Deduced from eqn. (8) in Reference 8.

Consider the locus of one of the component vectors, z_3 say, with respect to the origin O_3, Fig. 2:

$$z_3 = x_3 + jy_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

$$= -\frac{a_3}{1 + (\omega\tau_3)^2} + \frac{ja_3\omega\tau_3}{1 + (\omega\tau_3)^2} \quad . \quad . \quad . \quad (35)$$

Therefore

$$\frac{y_3}{x_3} = \omega\tau_3 = \tan\psi_3 \quad . \quad . \quad . \quad . \quad (36)$$

From eqns. (35) and (36),

$$x_3^2 + a_3x_3 + y_3^2 = 0$$

or

$$(x_3 + a_3/2)^2 + y_3^2 = (a_3/2)^2 \quad . \quad . \quad . \quad (37)$$

Hence the locus of z_3 is a circle, centre $(-a_3/2, 0)$ and diameter a_3.

If ω varies from 0 to ∞ the locus lies along the semicircle shown by the solid line in Fig. 2.

Similarly the loci of z_1 and z_2 are semicircles with diameters a_1 and a_2 respectively. The complete form of the locus is obtained by vector addition according to eqn. (33) and this may be accomplished by the construction indicated in Figs. 2 and 3. The method can, of course, be extended to discharges having any number of delayed effects.

The generalized form of the equivalent circuit of the glow discharge tube can be derived from the following equation for $n - 1$ delayed effects:

$$Z = a_0 - \frac{a_1}{1 + j\omega\tau_1} - \frac{a_2}{1 + j\omega\tau_2} \cdots - \frac{a_n}{1 + j\omega\tau_n} \quad . \quad (38)$$

From eqns. (28) and (29),

$$a_0 = R + a_1 + a_2 \ldots a_n \quad . \quad . \quad . \quad . \quad (39)$$

From eqns. (38) and (39),

$$Z = R + \frac{a_1 j\omega\tau_1}{1 + j\omega\tau_1} + \frac{a_2 j\omega\tau_2}{1 + j\omega\tau_2} + \ldots + \frac{a_n j\omega\tau_n}{1 + j\omega\tau_n} \quad . \quad (40)$$

Defining $L_1 = a_1\tau_1$, $L_2 = a_2\tau_2$, etc.,

$$Z = R + \frac{a_1 j\omega L_1}{a_1 + j\omega L_1} + \frac{a_2 j\omega L_2}{a_2 + j\omega L_2} + \ldots + \frac{a_n j\omega L_n}{a_n + j\omega L_n} \quad . \quad (41)$$

The second and subsequent terms of eqn. (41) represent a parallel combination of L and a and the equivalent circuit, including the effective capacitance, takes the form shown in Fig. 15.

621.316.74

The Institution of Electrical Engineers
Paper No. 3190 M
Mar. 1960
©

THE PRINCIPLE OF REVERSED LAG APPLIED TO ON-OFF TEMPERATURE CONTROL

By H. SUTCLIFFE, M.A., Associate Member.

(*The paper was first received 6th August, and in revised form 14th October*, 1959.)

SUMMARY

Cyclic fluctuations of temperature in an enclosure with a simple on-off control system are discussed. It is shown that such a system can be improved by using two control elements completely immersed in the fluid of the enclosure, and therefore not subject to datum error. The output signal of one of the control elements is required to have a lagging phase angle, and a theoretical analysis is made in which the thermal frequency responses of various configurations are compared. Conclusions are reached about the relative merits of different configurations, and a description is given of the application of the theoretical principles to an experimental temperature-controlled enclosure.

LIST OF SYMBOLS

a = Degree of concavity of exponential structure.
A = Area of isothermal surface.
J_0, J_1 = Bessel functions.
k = Diffusivity, cm^2/sec.
l = Length of structure in direction of heat flow, cm.
r = Radial distance from centre of sphere or axis of cylinder, cm.
R = Radius of sphere or cylinder, cm.
T = Temperature.
U = Product $r\theta$ for sphere.
v = Signal output from temperature-sensitive element, e.g. volts.
V = Time vector representing fluctuating component of v, e.g. volts.
P_h = Heater power, watts.
x = Distance along flow lines in linear flow, cm.
γ = Propagation coefficient, cm^{-1}.
θ, θ_0 = Time vectors representing fluctuating components of temperature.
t = Time.
ϕ = Lagging phase angle introduced by thermal propagation.
ψ = Leading phase angle required for improvement of system.
ω = Angular frequency of sinusoidal fluctuations, rad/s.

(1) INTRODUCTION

A common and economical method of effecting the automatic control of the temperature of an enclosure is shown in schematic form in Fig. 1. The fluid is supposed to be well stirred so that the whole of the enclosure is at a temperature T. A temperature-sensitive element detects whether T is greater or less than the desired temperature T_0. If T is less than T_0 a mechanism, indicated in the Figure by the line and arrowhead, closes the switch and applies constant power to a heater. If T exceeds T_0 the switch opens and the enclosure cools, by loss of heat to its surroundings. The principal disadvantage arising from this simple system is illustrated in Fig. 2, which shows the waveforms of heater power, P_h, and temperature, T, when the system has reached its state of quasi-equilibrium. Cyclic oscillations occur, the frequency being such that the waveform of T is in anti-phase with the fundamental component of P_h. Higher

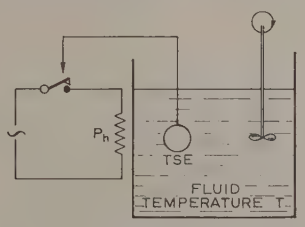


Fig. 1.—Schematic of enclosure with automatic temperature control
TSE. Temperature-sensitive element.

harmonics of P_h are greatly attenuated in the thermal transmission system and the T waveform is almost a perfect sinusoid. The deviations of T and $\bar{T}$ from T_0 are undesirable, and will be reduced proportionately by any modification to the system which results in a reduction in the amplitude of the sinusoidal fluctuation of T. Such a reduction would be obtained by supplying part of the heating power constantly and so reducing the magnitude of the fluctuating portion, but this solution

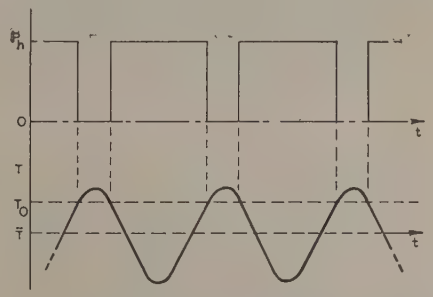


Fig. 2.—Waveforms of heating power P_h and temperature T in a system like that shown in Fig. 1.

requires a complicated mechanism if the range of control is not to be reduced. The paper will be restricted to the situation where P_h is applied by a single on-off switch, and improvements to the system can arise only from modifications which affect the instant of operation of the switch. More specifically, the natural frequency of the system requires to be increased so that the fluctuations of fluid temperature will be reduced because of the increased attenuation through the thermal transmission system.

The periodic characteristics of a typical system are illustrated in Fig. 3, which is a record of experimental results obtained with a small fluidized bed constructed for testing semiconductor devices. A waveform of P_h was available similar to that of Fig. 2 but having equal on and off times and with a choice of

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Sutcliffe is in the Department of Electrical Engineering, University of Bristol.

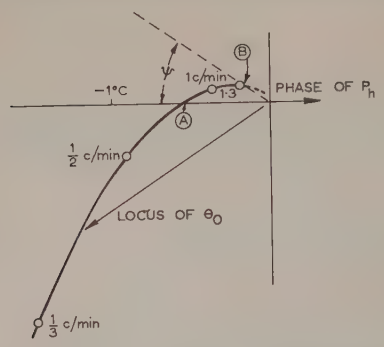


Fig. 3.—Frequency response of a small fluidized bed.
θ_0 represents the sinusoidal fluctuation in temperature.

several frequencies. The amplitude, and phase with respect to that of P_h, of the resulting fluctuations in bed temperature was measured at each of the frequencies. Results are represented by the time vector or phasor θ_0, whose locus on Fig. 3 shows how the fluctuations in temperature vary in amplitude and phase according to the frequency of operation of the system. A simple system, when the feedback loop is closed, will oscillate under conditions described by the point A on the locus, since at this point the phase of θ_0 lags that of P_h by 180°. A substantial reduction in amplitude of the fluctuation in temperature would be gained if the system could be made to oscillate at a frequency corresponding to a point such as B. This may be achieved by modifying the control system so as to introduce a leading phase angle which compensates for the lagging phase angle ψ of point B. Existing methods of introducing such a phase lead appear to fall into one of the following two categories:

(*a*) External phase-advance circuits or mechanisms.
(*b*) Addition of a lagging version of the P_h waveform.

Method (*a*) would be simple and economical if the temperature-sensitive element produced its signal in electrical form and if the frequency of oscillation were very much higher. A phase-advancing network of the class of which Fig. 4 is one example

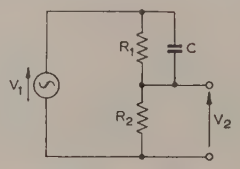


Fig. 4.—An electrical network giving a phase lead.

would be included in the loop. Unfortunately the impedance levels or component values prove to be impracticable for frequencies in the region of 1 cycle/min, and thus more complicated and expensive mechanisms are required.

Method (*b*) is used extensively, the simplest technique being the use of a small auxiliary heater round the temperature-sensitive element.[5] An alternative technique, used by the author, is to employ a second element in the vicinity of the heater and to add a much attenuated version of its output signal to the signal from the element immersed in the fluid. A third variation of the same general method is to use the P_h waveform to modify the effective datum level, T_0, cyclically, so that switching of heating power is advanced in phase along the waveform. A general criticism of method (*b*) is that datum errors are introduced owing to changes in heating power level or ambient temperature.

It was decided to investigate a third method, in which the disadvantages of (*a*) and (*b*) would be removed, by using temperature-sensitive elements totally immersed in the fluid and having a simple construction. The required output signal with phase advance would be obtained by using the principle of reversed lag, as described in the next Section.

(2) PRINCIPLE OF REVERSED LAG

It is instructive to redraw the network of Fig. 4 in the form shown in Fig. 5, by the use of Thévenin's theorem. It can be seen from Fig. 5 that the output signal V_2 has two components, the input signal V_1 and a signal V_3 which is dependent in amplitude and phase on the frequency of V_1. The signal V_3 is

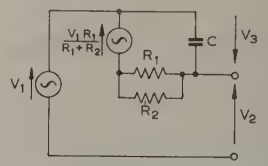


Fig. 5.—A rearrangement of Fig. 4 showing that phase lead arises from reversed phase lag.

a lagging version of a reversed fraction of V_1, so that the overall effect is to give V_2 a leading phase angle with respect to V_1.

The same principle governs the behaviour of the arrangement of temperature-sensitive elements shown in Fig. 6. The elements in this Figure have been drawn with a resemblance to thermocouples, but the Figure is intended to illustrate the general

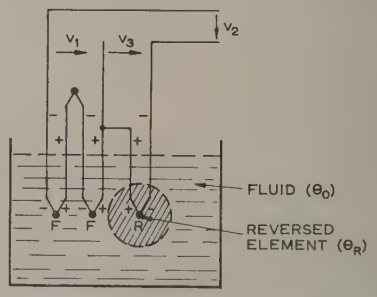


Fig. 6.—An arrangement of temperature-sensitive elements which provides an output signal whose fluctuations have a phase lead with respect to fluctuations of fluid temperature.
Element R has lagging phase angle and is connected in the reversed sense.

situation. The elements F, F are connected in the forward sense, and are in intimate contact with the fluid. The signal v_1 is then directly proportional to fluid temperature, and its fluctuating component is in phase with θ_0. The element R is enclosed within a structure which ensures that its mean temperature is that of the fluid, but thermal transmission through the structure introduces a phase lag into the fluctuating component of the temperature of the element. The sense of connection of R is reversed, and so the overall effect is that the fluctuating component of the output signal v_2 has a phase angle which leads that of θ_0. The mean value of v_2 is dependent entirely on the mean fluid temperature, so that the system has no inherent datum error. It is obvious, however, that there has been a sacrifice of sensitivity of detection of mean temperature, which might endanger the datum stability of the system. For this reason it is undesirable to have a system in which the magnitude of v_3 approaches that of v_1. It will be shown that this restriction imposes a limit to the phase lead which can be achieved.

Fig. 7 illustrates the relation between the vectors V_1, V_2 and V_3, which represent the fluctuating parts of the signals v_1, v_2 and v_3 and the vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_R$, which represent the fluctuating components of the temperature of the fluid and of the reversed element. It can be seen from the Figure that there are two

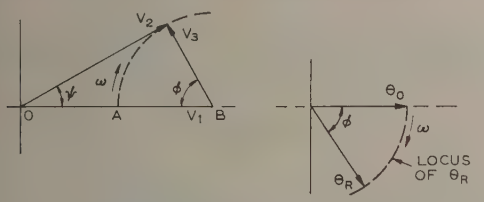


Fig. 7.—Vector relationship between fluctuating components of temperature and signals in Fig. 6.

requirements to be satisfied if the phase angle ψ is to be large. One of these is that length AB, which is the amplitude of V_3 at zero frequency, shall not be small compared with OB, the amplitude of V_1. The ratio illustrated in the Figure is AB = 0·5OB, which represents a sacrifice in stability of datum level by a factor of two. Any further increase in the ratio AB : OB would lead to serious loss of datum stability, since the controlling signal v_2 would become an increasingly small fraction of the signal v_1. The other requirement to be satisfied in order to obtain a large value of ψ concerns the shape of the locus of $\boldsymbol{\theta}_R$. The most favourable shape of locus is one for which the amplitude of $\boldsymbol{\theta}_R$ is subjected to the least attenuation as ϕ increases between 0 and 90°. A theoretical discussion of various configurations and their thermal frequency responses is given in the next Section.

(3) THEORETICAL COMPARISON BETWEEN THE THERMAL RESPONSES OF VARIOUS CONFIGURATIONS

It will be assumed that the heat flow is in a homogeneous material. It may be shown that the temperature at any point inside the material is determined by conditions at the boundary of the material and by the following relationship:[1,2]

$$\frac{\partial T}{\partial t} = k\nabla^2 T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where k is the ratio between the thermal conductivity and the volumetric specific heat.

The analysis in this Section will be restricted to situations where the temperature varies sinusoidally with time about a mean value. It is convenient to introduce the time vector or phasor $\boldsymbol{\theta}$ to represent the fluctuations of temperature, such that

$$T = \bar{T} + \mathscr{R}\boldsymbol{\theta} \exp j\omega t$$

Eqn. (1) then becomes

$$\nabla^2\boldsymbol{\theta} = \gamma^2\boldsymbol{\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $\gamma = (j\omega/k)^{1/2}$.

Relation (2) will now be used to find the thermal responses of a number of particular configurations.

(3.1) Linear Flow. Response Type L1

Fig. 8 shows one method of obtaining the L1 type of thermal response. Heat flows along the conducting rod and eqn. (2) takes the form:

$$\frac{\partial^2\boldsymbol{\theta}}{\partial x^2} = \gamma^2\boldsymbol{\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The left-hand end of the rod is in intimate contact with the fluid, and it will be assumed that the heat transfer from fluid

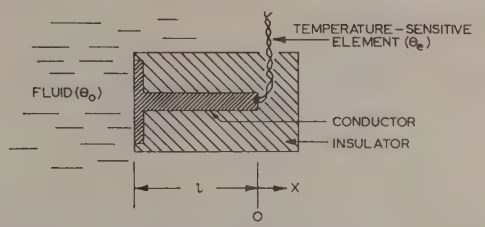


Fig. 8.—Configuration giving response type L1. Linear flow.

to rod is sufficiently effective to give the boundary condition $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ at $x = -l$. The other boundary condition is that the temperature gradient is zero at $x = 0$. The temperature at any point in the rod is then given by

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 \frac{\cosh \gamma x}{\cosh \gamma l} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The temperature at the end of the rod, where the temperature-sensitive element is located, is given by

$$\boldsymbol{\theta}_e = \boldsymbol{\theta}_0/\cosh \gamma l \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It is convenient for the purposes of comparison and design to obtain a plot of the vector locus $\boldsymbol{\theta}_e/\boldsymbol{\theta}_0$ as a function of the dimensionless parameter $|\gamma l| = l(\omega/k)^{1/2}$. This can be done using published tables,[3] and curve L1 in Fig. 17 is obtained.

(3.2) Linear Flow. Response Type L2

Fig. 9 is an example of a structure which will give the L2 type of response. The expansion of the rod is used to operate a

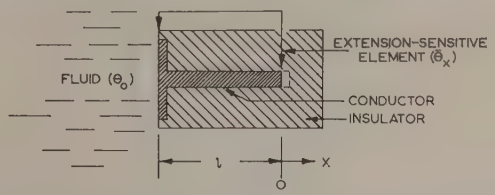


Fig. 9.—Configuration giving response type L2. Linear flow.

switch contact, for example, and so an expression is required for the mean temperature along the rod. From eqn. (4)

$$\bar{\boldsymbol{\theta}}_x = \frac{1}{l}\int_{-l}^{0} \boldsymbol{\theta}_0 \frac{\cosh(\gamma x)}{\cosh(\gamma l)} dx$$

$$\bar{\boldsymbol{\theta}}_x = \boldsymbol{\theta}_0 \frac{\tanh(\gamma l)}{\gamma l} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The ratio $\bar{\boldsymbol{\theta}}_x/\boldsymbol{\theta}_0$ is given in graphical form by the curve L2 in Fig. 17. As in the previous case the tables of Reference 3 were used, and the results are given in terms of the parameter $|\gamma l|$.

(3.3) Linear Flow. Response Type M1

The M1 type of response will be discussed in order to demonstrate the undesirable effects which arise when a structure such as that of Fig. 1 is modified by extending the conductor past the temperature-sensitive element, or by adding a massive conductor at this point. The extreme case of such modifications will result in a reflectionless system of heat propagation, idealized in Fig. 10. The transmission system is matched (hence the M

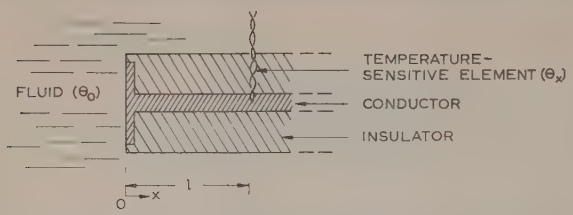


Fig. 10.—Configuration giving response type M1. Linear flow.

in the nomenclature) and the temperature at any point, derived from eqn. (3), is given by

$$\theta = \theta_0 \exp(-\gamma x) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and, in particular,

$$\theta_x = \theta_0 \exp(-\gamma l) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The locus of θ_x/θ_0 is shown as curve M1 in Fig. 17 with $|\gamma l|$ as parameter.

(3.4) Linear Flow. Response Type M2

A matched system is again considered, but now the sensing element is detecting the extension of a length of the rod as illustrated in Fig. 11. The extension is proportional to the mean

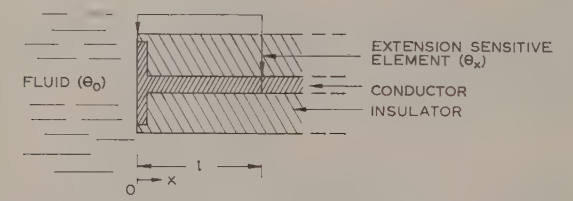


Fig. 11.—Configuration giving response type M2. Linear flow.

temperature of length l of the rod, and may be obtained from the following expressions using eqn. (7):

$$\bar{\theta}_x = \frac{1}{l}\int_0^l \theta_0 \exp(-\gamma x)dx$$

i.e.

$$\bar{\theta}_x = \theta_0 \frac{1 - \exp(-\gamma l)}{\gamma l} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The locus of $\bar{\theta}_x/\theta_0$ in terms of the parameter $|\gamma l|$ is illustrated in Fig. 17 by curve M2.

(3.5) Radial Flow, Cylindrical Structure, Response Type C1

The C1 type of response arises when a temperature-sensitive element is located on the axis of a lengthy conducting cylinder immersed in the fluid, as illustrated in Fig. 12. The structure

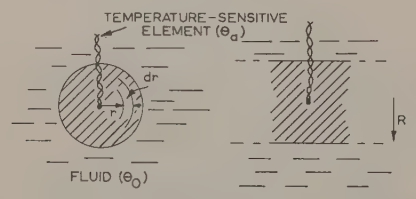


Fig. 12.—Cylindrical configuration giving response type C1. Radial flow.

has radial symmetry, and eqn. (2) may be expressed in terms of radial distance only.

$$\frac{\partial^2\theta}{\partial r^2} + \frac{1}{r}\frac{\partial\theta}{\partial r} - \gamma^2\theta = 0 \quad . \quad . \quad . \quad . \quad (10)$$

If, in eqn. (10), γ^2 were real, the solution would be expressed in terms of modified Bessel functions, but since, in fact, γ^2 is complex, it seems to be more convenient to follow the nomenclature of McLachlan[4] and to express the solution as follows:

$$\theta = AJ_0(j^{3/2}|\gamma r|) + BK_0(j^{3/2}|\gamma r|) \quad . \quad . \quad . \quad (11)$$

where J_0 and K_0 are modified Bessel functions.

θ is finite at $r = 0$, and since $K_0(0)$ is infinite it follows that $B = 0$; $\theta = \theta_0$ at $r = R$, and so eqn. (11) becomes

$$\theta = \theta_0 \frac{J_0(j^{3/2}|\gamma r|)}{J_0(j^{3/2}|\gamma R|)} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

The temperature-sensitive element is located on the axis, and at this point [since $J_0(0) = 1$] the temperature is

$$\theta_a = \theta_0 / J_0(j^{3/2}|\gamma R|) \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The locus θ_a/θ_0 is required, and this is obtained by using tables of J_0 given in Reference 4 in polar form. The locus is plotted in Fig. 17 in terms of the parameter $|\gamma R|$.

(3.6) Radial Flow, Cylindrical Structure, Response Type C2

The axial extension of a long cylindrical rod immersed in the fluid gives the C2 type of response. The configuration is shown in Fig. 13, and it can be shown that the extension will be proportional to the mean temperature, the average being taken over the cross-sectional area of the rod. This average is obtained using eqn. (12).

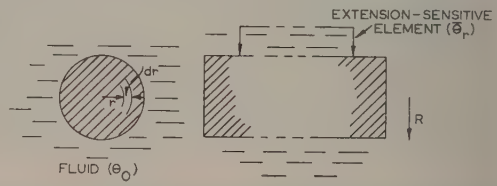


Fig. 13.—Cylindrical configuration giving response type C2. Radial flow.

$$\bar{\theta}_r = \frac{1}{\pi R^2}\int_0^R \frac{\theta_0 J_0(j^{3/2}|\gamma r|)}{J_0(j^{3/2}|\gamma R|)} 2\pi r dr \quad . \quad . \quad . \quad (14)$$

$$= \theta_0 \frac{2}{j^{3/2}|\gamma R|} \frac{J_1(j^{3/2}|\gamma R|)}{J_0(j^{3/2}|\gamma R|)} \quad . \quad . \quad . \quad . \quad (15)$$

The ratio $\bar{\theta}_r/\theta_0$ is plotted as a function of the parameter $|\gamma R|$ to give curve C2 in Fig. 17. Tables of Bessel functions in polar form from Reference 4 were used to draw the curve, but it is of interest to note that a similar equation represents the ratio of a.c. internal admittance to d.c. conductance of round straight wire. The corresponding impedance ratio is shown in tabular or graphical form in many texts, and would be used by substituting $1/k$ for the product μG (permeability times electrical conductance).

(3.7) Radial Flow, Spherical Structure, Response Type S1

A temperature-sensitive element located at the centre of a sphere immersed in the fluid will give the S1 type of response. The structure is shown in Fig. 14. The flow will have spherical symmetry and so the appropriate forms of eqns. (1) and (2) are as follows:

$$\frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial r^2} + \frac{2}{r}\frac{\partial T}{\partial r}\right) \quad . \quad . \quad . \quad . \quad (16)$$

$$\frac{\partial^2\theta}{\partial r^2} + \frac{2}{r}\frac{\partial\theta}{\partial r} = \frac{j\omega\theta}{k} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

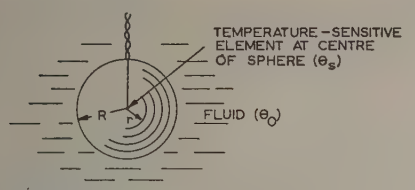


Fig. 14.—Spherical configuration giving response type S1. Radial flow.

The analysis is simplified by making the substitution $U = r\theta$, where U, like θ, is a time vector or phasor. Substituting in eqn. (17) gives

$$\frac{\partial^2 U}{\partial r^2} = \frac{j\omega}{k} U = \gamma^2 U \quad . \quad . \quad . \quad . \quad . \quad (18)$$

i.e. $$U = A \cosh(\gamma r) + B \sinh(\gamma r) \quad . \quad . \quad . \quad (19)$$

The boundary conditions are $U = 0$ when $r = 0$
$U = R\theta_0$ when $r = R$

From eqn. (19) $$U = r\theta = R\theta_0 \frac{\sinh \gamma r}{\sinh \gamma R}$$

i.e. $$\theta = \theta_0 \frac{R}{r} \frac{\sinh \gamma r}{\sinh \gamma R} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

At the centre of the sphere, $\theta = \theta_s$, $r \to 0$, $\sinh \gamma r \to \gamma r$; which gives an expression for the temperature fluctuation:

$$\theta_s = \theta_0 \frac{\gamma R}{\sinh \gamma R} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Curve S1 in Fig. 17 is the locus of θ_s/θ_0 in terms of the parameter $|\gamma R|$.

(3.8) Radial Flow, Spherical Structure, Response Type S2

The S2 type of response arises when a sensing element measures the volumetric expansion of a liquid contained in a rigid thin-walled spherical container of high conductivity immersed in the fluid. Fig. 15 shows the arrangement, which is, of course, an

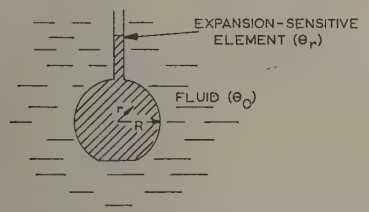


Fig. 15.—Spherical configuration giving response type S2. Radial flow.

idealized version of the common thermometer. The sensing element detects the volumetric expansion, which is proportional to the mean temperature $\bar{\theta}_r$ over the volume of the sphere.

$$\bar{\theta}_r = \frac{\int_0^R 4\pi r^2 \theta dr}{\frac{4}{3}\pi R^3} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

From eqns. (20) and (22), by simple integration, the following expression for $\bar{\theta}_r$ is obtained:

$$\bar{\theta}_r = \theta_0 3[1/\gamma R \tanh \gamma R - 1/(\gamma R)^2] \quad . \quad . \quad . \quad (23)$$

The locus of $\bar{\theta}_r/\theta_0$ is plotted as curve S2 in Fig. 17 in terms of the parameter $|\gamma R|$. Values of $\tanh \gamma R$ were taken from the tables in Reference 3.

(3.9) Highly Convergent Flow. Exponential Structure. Optimum Response

Curves L1, C1 and S1 show increasingly advantageous properties in that curve S1 exhibits least attenuation for a given phase angle. An explanation for this effect may be given by considering the three equivalent electrical circuit analogues, each of which consists of a multi-section ladder network of series R and parallel C elements. The sections in the linear-flow analogue are all identical, while in the cylindrical-flow and spherical-flow analogues the impedance of the network components increases towards the terminals which represent conditions at the centre of the cylinder or sphere, the increase in impedance being more pronounced for the sphere. The effect of this increase in impedance is to reduce the attenuation in any particular section for a given change of phase, and so to produce a more favourable overall response. This suggests that a network should be considered in which each succeeding section is of such relatively high impedance that its loading effect on the preceding section is negligible. Each section then contributes a small change of phase with negligible attenuation, and the overall response locus is a circle with the origin as centre.

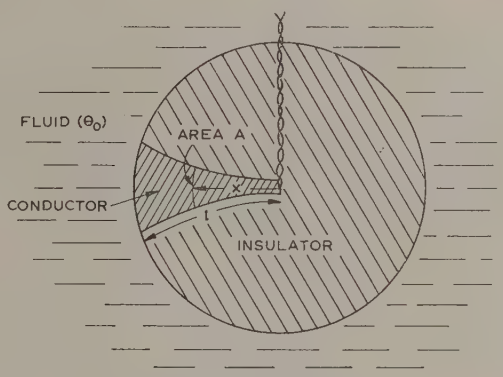


Fig. 16.—Exponentially tapered conductor giving response type E. Highly convergent flow.

The considerations suggest that a structure of the type shown in Fig. 16 will be capable of producing a response whose vector locus lies outside that of the spherical structure.

An exact analysis is difficult, but if it is assumed that the temperature gradient is nearly uniform over an isothermal surface, eqn. (2) may be written in the form

$$\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial \theta}{\partial x} \frac{\partial A}{\partial x} \frac{1}{A} = \gamma^2 \theta \quad . \quad . \quad . \quad . \quad (24)$$

where x is distance measured along the flow lines and A is the area of the isothermal surface.

Eqn. (24) becomes tractable if A increases exponentially with distance. It will therefore be assumed that $A = A_0 \exp ax/l$, where the parameter a is a measure of the concavity of the profile of the conductor. Eqn. (24) now becomes

$$\frac{\partial^2 \theta}{\partial x^2} + \frac{a}{l} \frac{\partial \theta}{\partial x} - \gamma^2 \theta = 0 \quad . \quad . \quad . \quad . \quad (25)$$

i.e. $$\theta = M \exp\left(\left\{-\frac{a}{2l} + \sqrt{\left[\left(\frac{a}{2l}\right)^2 + \gamma^2\right]}\right\}x\right)$$
$$+ N \exp\left(\left\{-\frac{a}{2l} - \sqrt{\left[\left(\frac{a}{2l}\right)^2 + \gamma^2\right]}\right\}x\right) \quad . \quad (26)$$

The arbitrary constants M and N are found by inserting the following boundary conditions:

(i) $\dfrac{\partial\theta}{\partial x} = 0$ when $x = 0$

(ii) $\theta = \theta_0$ when $x = l$

Then
$$N = M\frac{-\frac{a}{2l} + \sqrt{\left[\left(\frac{a}{2l}\right)^2 + \gamma^2\right]}}{\frac{a}{2} + \sqrt{\left[\left(\frac{a}{2l}\right)^2 + \gamma^2\right]}} \quad . \quad . \quad . \quad (27)$$

from condition (i) and eqn. (26).

The artificial case where $\left(\dfrac{a}{2l}\right)^2 \gg |\gamma|^2$ will be considered first. The second term in eqn. (26) may be neglected, and condition (ii) and eqn. (26) give, at $x = 0$,

$$\theta_1 = \theta_0 \exp\left\{\frac{a}{2} - \sqrt{\left[\left(\frac{a}{2}\right)^2 + (\gamma l)^2\right]}\right\} \quad . \quad . \quad . \quad . \quad (28)$$

$$= \theta_0 \exp\left(\frac{a}{2}\left\{1 - \left[1 + \frac{1}{2}\frac{(\gamma l)^2}{(a/2)^2} + \ldots\right]\right\}\right)$$

$$\simeq \theta_0 \exp\left[-(\gamma l)^2/a\right] = \theta_0 \exp\left[-j|\gamma l|^2/a\right] \quad . \quad (29)$$

Eqn. (29) shows that the locus θ_1/θ_0 becomes a circle about the origin, provided that a is sufficiently large. We wish to know how closely this optimum locus can be approached in reality, and to this end it will be assumed that $a = 8$ is the greatest value which may be used. This figure is chosen somewhat arbitrarily, but arises from the difficulties of making a structure like Fig. 9 with a temperature-sensitive element at the narrow end. The value $a = 8$ corresponds to an increase of 54 times in diameter and an increase of 3 000 times in area.

It is not intended to plot the whole locus, since it is know that this lies between curve S1 and the circle, but it is of grea interest to find a value of θ_1/θ_0 for which the phase angle is i

Table 1

Curve	Fig. No.	Function	Parameter
L1	8	$\dfrac{1}{\cosh\ \gamma l}$ where $\gamma = (j\omega/k)^{1/2}$	$\|\gamma l\|$
L2	9	$\dfrac{\tanh\ \gamma l}{\gamma l}$	$\|\gamma l\|$
M1	10	$\exp(-\gamma l)$	$\|\gamma l\|$
M2	11	$\dfrac{1 - \exp(-\gamma l)}{\gamma l}$	$\|\gamma l\|$
C1	12	$\dfrac{1}{J_0(j^{3/2}\|\gamma R\|)}$	$\|\gamma R\|$
C2	13	$\dfrac{2J_1(j^{3/2}\|\gamma R\|)}{j^{3/2}\|\gamma R\|J_0(j^{3/2}\|\gamma R\|)}$	$\|\gamma R\|$
S1	14	$\dfrac{\gamma R}{\sinh\ \gamma R}$	$\|\gamma R\|$
S2	15	$\dfrac{3}{\gamma R \tanh\ \gamma R} - \dfrac{3}{(\gamma R)^2}$	$\|\gamma R\|$
E	16	See eqns. (26) and (27)	

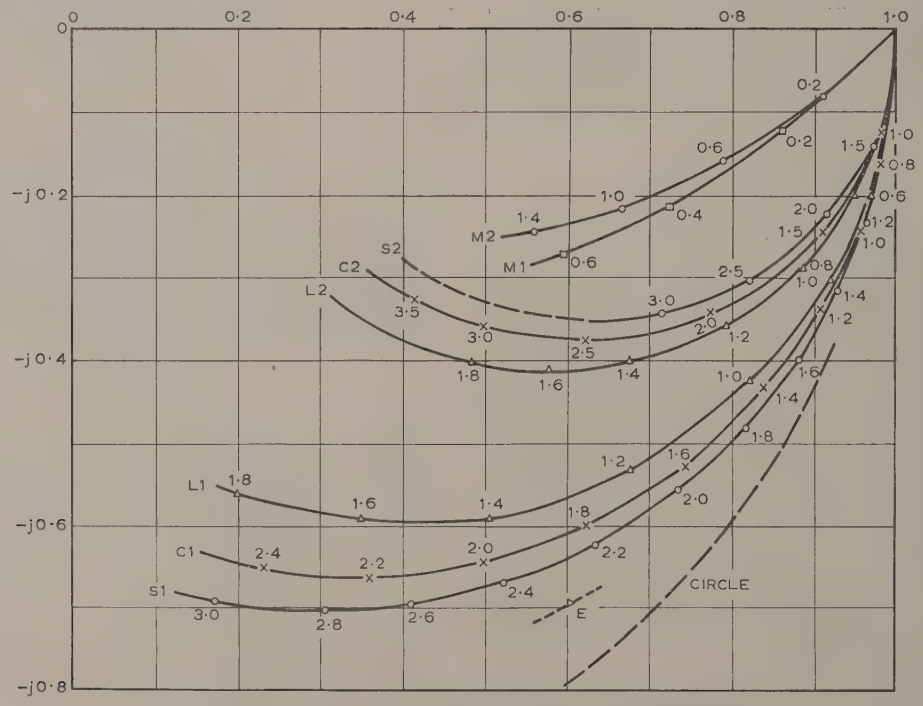


Fig. 17.—Thermal frequency responses of various configurations (see Table 1).

he region of 60° lagging. The position of this single point n Fig. 17 will be adequate for the purpose of deciding whether he improvements obtainable from structures of the type of ig. 16 are really worth the extra complication of construction. We wish to choose, then, a value of $|\gamma l|$ which will produce a alue of θ_1/θ_0 with the desired angle. It may be assumed that qn. (29) will be approximately correct for the purpose of hoosing a useful value, and so we arrive at the following umerical values for substitution in eqns. (26) and (27):

$$a = 8,\ |\gamma l| = 8, \text{ from which } \theta_1 = \theta_0 0{\cdot}92 \angle -49°$$

The point E on Fig. 17 corresponds to this result, and is seen o lie not quite half way between curve S1 and the circle.

Examination of Fig. 17 shows that the various loci fall into hree groups. Curves M1 and M2, which relate to linear heat low under matched conditions, are clearly unsuitable, and struc- ures giving this type of response are incapable of providing an ppreciable phase lead.

Curves L, C2 and S2 form a group with more favourable roperties, and are worth consideration where the application nvolves little in the way of mechanical complications and where moderate phase lead is adequate. An example of this situation s shown in Fig. 18. A bimetal strip in intimate contact with the luid is required to operate a pair of switch contacts. The notion must therefore be transmitted through a flexible portion f a protective enclosure, and it is convenient to introduce a hort length of bimetal strip connected in the reversed sense within the enclosure. Heating of the short reversed section is y lengthwise conduction, and the reversed lag is estimated from urve L2. A model based on Fig. 18 was constructed and

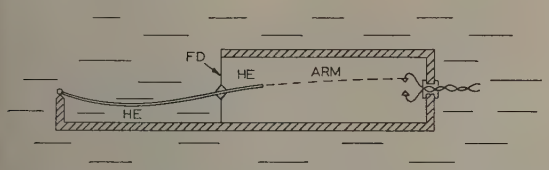


Fig. 18.—A temperature controlling element with L2 type of reversed lag.

FD. Flexible diaphragm.
HE. High-expansion side of bimetal strip.

erved to illustrate the principle of reversed lag. This particular ype of structure has not been investigated in great detail, since t is not suitable in the miniature sizes required in applications vhich were of particular interest to the author. It also suffers rom the effect of fluid turbulence, which causes chatter at the witch contacts.

Curves L1, C1 and S1 form a group in which the amplitude f the response is well maintained as the phase angle lags. Useful values of phase lead ψ should be obtainable, and that his is indeed the case is illustrated by a practical application lescribed in the next Section.

(4) A PARTICULAR EXAMPLE OF THE PRINCIPLE OF REVERSED LAG

A small fluidized bed was found to have the frequency response hown in Fig. 3 when its temperature was 40° C above ambient. Changes in mean temperature of the bed made little change in he shape of the locus as a function of frequency but simply ltered the scale. Thus, for a mean temperature 80° C above mbient, the magnitude of θ_0 at point A was increased from)·5° C (peak to peak) to 1° C. Such fluctuations were regarded as excessive for the applications with which the bed was to be associated, and it was decided to use the principle of reversed lag to increase the oscillation frequency to the region 1·3 cycles/min, and so to reduce the amplitude of the fluctuations to rather less than one half their former value. The forward and reverse temperature-sensitive elements were both thermocouples, and a voltage divider across the reversed element was used to vary and adjust the amount of reversed signal, i.e. the ratio AB/OB in Fig. 7. The total output signal was compared with a potential difference whose value determined the datum temperature, a chopper-type differential amplifier of high gain being used to determine the sense of the difference. The output signal of the amplifier drove the coil of an electromagnetic relay whose contacts switched power to the bed heater.

A lagging thermal response of the required type was determined as follows. Curve L1 in Fig. 17 shows that a suitable value of the parameter $|\gamma l|$ is in the region 1·1, but, in view of the difficulty of controlling heat flow, it is desirable that the value of l is adjustable in preliminary measurements. Such an adjustment may be obtained by using heat flow along the thermocouple itself, and by varying the length covered by a thermal insulator. Considerations of the space available showed that $l = 1$ cm would be a suitable length, and so the following calculations were made in order to find a material whose diffusivity has the desired magnitude:

$$|\gamma l|^2 = \omega l^2/k = 1{\cdot}21,\ \omega = 2\pi 1{\cdot}3/60 \text{ rad/s}$$
$$l = 1 \text{ cm}$$
$$k = 0{\cdot}11 \text{ (C.G.S. units)}$$

Tables of thermal properties of materials in Reference 1 give the diffusivity of steel and cast iron as 0·12, and so it was decided to use iron-constantan as the thermocouple. Heat flow along the constantan lead was made negligible by using very fine wire (No. 38 s.w.g.) while the iron lead was of 0·020 in-diameter piano wire. An adjustment of the length l was made by drilling a hole in a cork and sliding the cork along the thermocouple. The system proved to be satisfactory, and the expected improvement in fluctuation amplitude was obtained using a value of l given by theoretical considerations. Thus the experimental results, though not exhaustive, support the argument that the theory provides a satisfactory basis for the design of particular control systems.

(5) CONCLUSION

It has been shown that the principle of reversed lag can be applied to an on-off temperature-control system with useful but not spectacular reductions in the amplitude of the cyclic fluctuation in temperature. Design data for the temperature-sensitive elements may be derived from Fig. 17 in conjunction with a measurement of the frequency response of the temperature deviations of the controlled enclosure.

(6) REFERENCES

(1) CARSLAW, H. S., and JAEGER, J. C.: 'Conduction of Heat in Solids' (Oxford University Press, 1947).
(2) CARSLAW, H. S., and JAEGER, J. C.: 'Operational Methods in Applied Mathematics' (Oxford University Press, 1949).
(3) KENNELLY, A. E.: 'Tables of Complex Hyperbolic and Circular Functions' (Harvard University Press, 1921).
(4) MCLACHLAN, N. W.: 'Bessel Functions for Engineers' (Oxford University Press, 1934).
(5) KLAPP, J. K.: 'Notes on the Design of Temperature Control Units', *General Radio Experimenter*, 1944, **19**, No. 3, p. 1.

621.387.422: 621.317.79

The Institution of Electrical Engineers
Paper No. 3207 M
Mar. 196

©

THE A.C. OPERATION OF ION CHAMBERS AT HIGH FLUXES

By J. WATSON, B.Sc., S.M., Ph.D., Graduate.

(*The paper was first received* 15*th June, and in revised form* 19*th November*, 1959.)

SUMMARY

When a sinusoidally varying p.d. is applied between the electrodes of an ion chamber subject to ionizing radiations, the motion of the charge carriers and the form of the current in the external circuit depend upon the applied voltage and frequency, the gas filling and the flux level.

A brief analysis of such conditions is made with space-charge effects neglected and is then modified semi-empirically to take account of space-charge phenomena. In the course of this analysis a definition of 'saturation' as applied to chambers using alternating polarizing voltages is proposed, and a series of experiments is described showing how variations in the aforementioned parameters affect this condition. Throughout the tests a standard A.E.R.E. RC1 ion chamber was used and was irradiated in the reactors Gleep and Bepo.

(1) INTRODUCTION

To date, the instrumentation and control channels associated with the higher flux levels encountered in nuclear reactors have been based upon the use of the d.c. polarized ion chamber as the initial transducer. This has resulted in the inevitable utilization of the d.c. amplifier with its associated stabilizing circuits, most of which involve a mechanical chopper at some point. Such a mechanical device is not considered desirable from a reliability aspect, and this has led to the suggestion that an alternating voltage be applied to the chamber itself, and a standard a.c. amplifier employed thereafter.

The application of an alternating voltage across the chamber means that a capacitive current must flow in addition to the ionization current. This component may be balanced out by a simple backing-off condenser, as shown in Fig. 1, when the

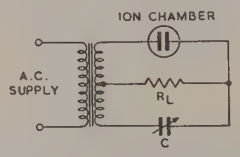


Fig. 1.—Basic backing-off circuit.

remaining current will develop a voltage across the resistor R_L. This has been done by two workers,[1,2] one of whom used a variable condenser in the backing-off arm, while the other favoured an identical, but unirradiated ion chamber. Unfortunately, both systems tended to distort the waveform of the output voltage by reason of their poor frequency response—a situation which also gave rise to an inaccurate evaluation of the output voltage/frequency curve. This poor frequency response was engendered by the time-constant of the load resistance, R_L, in conjunction with the stray capacitance across it, C_S, due to the screened cable to the chamber, which caused the response of both systems to drop to the −3 dB point at about 60 c/s.

The apparatus here employed used a virtual-earth amplifier following the junction of the chamber and balance condenser rather than a load resistor, and the extremely low input impedance presented effectively swamped the reactance of C_S and led to a frequency response cutting off only at about 5 kc/s. Moreover, the ionization component of the current was found to be of the same order as the capacitive component when a standard RC1 ion chamber was inserted into an experimental hole in the reactor Gleep and an alternating voltage at 70 c/s was applied, thus making unnecessary the Wagner earth which was imperative in the low-flux laboratory systems. This arrangement was supplied from a signal generator via an accurate phase-splitting amplifier having a variable output voltage. Thus, after balancing out the purely capacitive component of the output voltage with the chamber unirradiated, the waveform of the ionization component could be displayed and measured as the chamber was inserted into the pile.

A complete block diagram of the apparatus is given in Fig. 2.

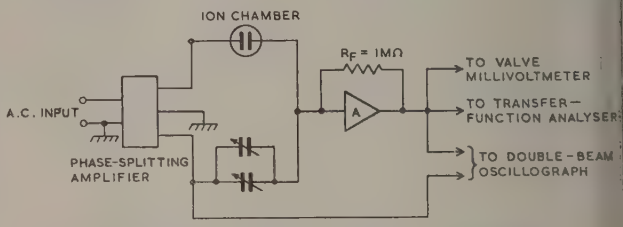


Fig. 2.—Block diagram of experimental apparatus.

from which it will be seen that, apart from the waveform monitor, a valve voltmeter and transfer-function analyser were connected to the output point of the virtual-earth amplifier.

From standard theory it can be shown that the voltage appearing at this output is given by IR_F, where R_F is the feedback resistance and I the input current. The supply at 70 c/s was chosen as a compromise avoiding a frequency so low as to cause beating with the mains and so high as to mask the typical waveforms due to ion motion, as described below. A further facility was the provision of a standard d.c. channel to which the chamber could be coupled in order to make direct comparisons with the a.c. case. The essential geometry of the chamber is shown in Fig. 3.

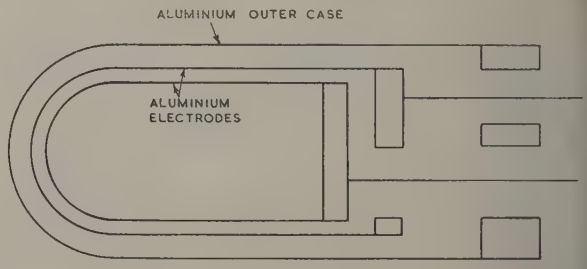


Fig. 3.—RC1 electrode configuration, showing relative positions of aluminium electrodes and airtight outer casing.
Insulation and other details omitted.
Overall length = 8¾ in.
Overall diameter = 3½ in.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Watson is with Ericsson Telephones, Ltd.

(2) THEORY OF OPERATION OF THE SATURATED D.C. ION CHAMBER

When ions are produced at a constant rate in the presence of an electric field, three processes combine to remove these ions, namely recombination, diffusion to the walls and collection by the field. Considering for the moment an electro-negative gas in which the electrons quickly become attached to neutral atoms, thus forming negative ions), the volume recombination rate is obviously proportional to the concentrations of the positive and negative ions, i.e.

$$\frac{-dn^+}{dt} = \frac{-dn^-}{dt} = \rho n^+ n^- \quad . \quad . \quad . \quad . \quad (1)$$

where ρ is the recombination coefficient[3] and n the concentration of charge carriers.

If $n^+ = n^-$,

$$\frac{-dn}{dt} = \rho n^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The loss of ions by diffusion out of a small element of volume is given by

$$\frac{dn}{dt} = D\nabla^2 n$$

or, in the one-dimensional case—as is applicable in an ion chamber having a small plate spacing—

$$\frac{dn}{dt} = D\frac{d^2n}{dx^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where D is the diffusion coefficient.

The effect of the applied electric field is to cause the ions to drift along it at a velocity $E\mu$, where E is the field strength and μ the mobility. In a parallel-plate ion chamber, where the distance, d, between the plates is much smaller than the length or breadth of the plates, $E = V_{dc}/d$, where V_{dc} is an applied direct voltage. The uniformity of field implied by this expression can be so only if space-charge effects are neglected—a condition which does not occur in practice, but which is a useful assumption for the purposes of elementary theory.

Thus, the transit time, T, or the time taken for an ion to traverse the distance, d, between the plates will be given by

$$T = \frac{d}{E\mu} = \frac{d^2}{V_{dc}\mu} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Such an ion traversing the active volume will be collected only if it does not suffer recombination on the way or is not carried by diffusion processes into some part of the chamber other than the collecting plate and is there neutralized. In other words, the transit time must be shorter than the recombination or diffusion times if the probability of collection is to be high. Comparison of such times will not be entered into here, since they have been fully dealt with in the standard literature,[4] and it is well known that such conditions can be achieved by a suitable choice of polarizing voltage.

This is the condition known as 'saturation', where the chamber-current/applied-voltage curve becomes a plateau and no further current rise can take place until sufficient energy is supplied to cause secondary ionization. The complete equations showing the balance between ion production, recombination, diffusion and collection are notoriously intractable,[5] especially when space-charge effects are taken into account, and so a semi-empirical approach will be followed when investigating the case of an alternating applied voltage.

(3) THEORY OF OPERATION OF THE SATURATED A.C. ION CHAMBER

(3.1) Instantaneous Chamber Current

When an ion is in motion in the field between two electrodes, the changing value of induced charge on these electrodes causes a current to flow in the external circuit. This means that ion collection is not necessary for the production of such a current. Mathematically, equating the work done on the charged ion by the field to the work done by the external current gives $Eeds = v_{ac}idt$, where ds is the small element of distance traversed by the charge, e, in the time dt, and i is the instantaneous current, given by

$$i = \frac{Ee}{v_{ac}}\frac{ds}{dt}$$

where v_{ac} is the instantaneous applied voltage
But since ds/dt = velocity = $\mu v_{ac}/d$,

$$i = \frac{\mu v_{ac} e}{d^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Thus the total instantaneous current for a number of ions of one sign in a chamber of volume v at any instant is

$$i' = \frac{nv\mu v_{ac} e}{d^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Since v, e, μ and d are constants, a convenient set of 'chamber constants' can be defined, relating to positive ions negative ions or electrons for any given chamber at any given gas filling at a known pressure, namely

$$C^+ = \frac{ve\mu^+}{d^2} \text{ for positive ions} \quad . \quad . \quad . \quad (7a)$$

$$C^- = \frac{ve\mu^-}{d^2} \text{ for negative ions} \quad . \quad . \quad . \quad (7b)$$

$$C_e = \frac{ve\mu_e}{d^2} \text{ for electrons} \quad . \quad . \quad . \quad . \quad . \quad (7c)$$

Thus, for any chamber, eqn. (6) becomes

$$i' = Cnv_{ac} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

(3.2) Output Waveform of the Chamber

When an alternating voltage is applied to the electrodes of a chamber, the resulting ion-current waveshapes will not be sinusoidal, but will depend upon the behaviour of the ions and electrons in the alternating field. If the effects of space charge are ignored, the field strength in the active volume will be proportional to the applied voltage. Therefore, at the beginning of a voltage cycle, the drift velocity of the ions will increase with the voltage, as will the rate of collection. Provided that the voltage rises to a sufficiently high level, some point will be reached where the rate of collection is equal to the rate of generation, and saturation as previously defined will set in. Any increase in voltage above this value will serve only to reduce the total number of ions present in the active volume, and the ion current will not vary.

When the voltage falls below this saturation level at the end of a half-cycle, the ion current will fall and the concentration will increase rapidly. When the voltage reverses, therefore, a large number of ions will be present in the active volume, and their movement (now in the opposite direction) will constitute a peak in the current waveform. This chamber-current waveform may therefore be expected to have an initial peak followed by a

saturation plateau for each type of charge carrier. Furthermore, if the filling is a free-electron gas, the initial transient due to the electrons may be expected to terminate long before the heavy-ion peak has formed, owing to the much higher mobilities involved.

If the voltage does not rise sufficiently high for saturation to occur, some of the ions formed will never be collected but will oscillate between the plates until removed by recombination or diffusion processes. Their movement will nevertheless constitute a current in the external circuit whose magnitude will be dependent upon the product of the ion concentration eventually reached and the voltage applied.

Using eqn. (8), the form of the current wave produced by a saturated chamber can be determined as follows, neglecting space-charge effects. In Fig. 4, curve (*a*) represents a sinusoidal

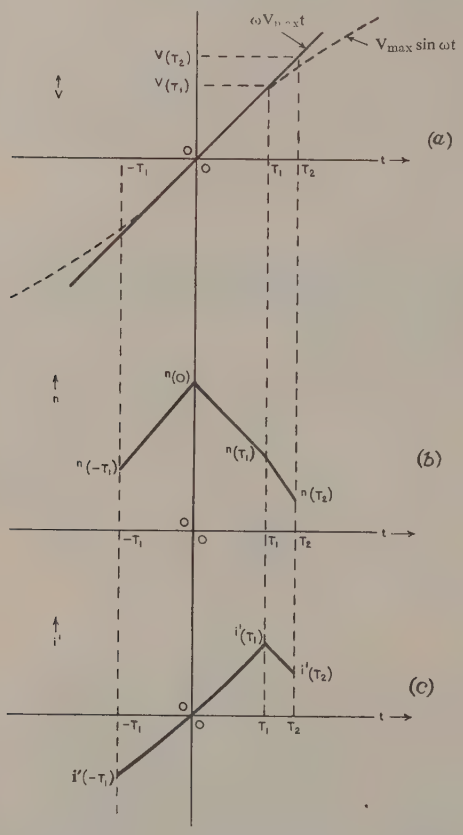


Fig. 4.—Idealized variation of applied voltage, ion concentration and ion current with time.

voltage applied to a parallel-plate chamber. The amplitude and frequency of this voltage are assumed to be such that all the ions formed during one half-period can be collected before the end of the next, and that negligible recombination or loss by diffusion occurs. If, for example, a positive ion is formed at time $t = 0$ very near the positive-going electrode, it will take one transit time, T, to cross to the negative-going plate.

From eqn. (4) T is given by

$$T = \frac{d^2}{V_{av}\mu^+} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where V_{av} is the average value of v_{ac} over the period from $t = 0$ to $t = T_1$. This is given by

$$V_{av} = \frac{V_{max}}{T_1} \int_0^{T_1} \sin \omega t dt$$

which, when inserted into eqn. (9), gives

$$T_1 = \frac{1}{\omega} \text{ arc cos } \left(1 - \frac{\omega d^2}{\mu^+ V_{max}}\right) \quad . \quad . \quad . \quad (10)$$

Similarly, the transit time $T_2 - T_1$ starting at $t = T_1$ will be given by

$$T_2 = \frac{1}{\omega} \text{ arc cos } \left(1 - \frac{2\omega d^2}{\mu^+ V_{max}}\right) \quad . \quad . \quad . \quad (11)$$

and so on.

Consider the time $t = -T_1$, and at this time let the ion concentration be $n_{(-T_1)}$. Then, by definition, all of the group of $n_{(-T_1)}$ ions will have been collected when $t = 0$. But during the period from $t = -T_1$ to $t = 0$ a further group of ions will have been formed, given by ST_1, where S is the (constant) rate of production. In order to determine the proportion of ions collected during transit times occurring near the origin, the voltage wave can be assumed to approximate to a ramp function having the same slope as the sine wave at $t = 0$. The slope being ωV_{max} for the sine wave $V_{max} \sin \omega t$, the ramp voltage at any time is now

$$V_{(t)} = \omega V_{max} t \quad . \quad . \quad . \quad . \quad . \quad (12)$$

It is shown in Section 9 that, for a ramp voltage which is falling, one third of the ions formed during a transit time are collected by the end of that transit time. Consequently, of the group of ST_1 ions generated during the interval from $t = -T_1$ to $t = 0$, only $ST_1/3$ are collected when $t = 0$.

Therefore $$n_{(0)} = \tfrac{2}{3}ST_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

It is also shown that, when the voltage ramp is rising, two thirds of the ions formed during a transit time are collected by the end of that transit time.

Hence $$n_{(+T_1)} = \tfrac{1}{3}ST_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

In order to obtain the actual values of T_1 and T_2 under these circumstances, the procedure culminating in eqns. (10) and (11) may be followed, but using the ramp voltage in place of the sinusoidal voltage. It will be found that

$$T_1 = d\sqrt{\frac{2}{\mu\omega V_{max}}} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and $$T_2 = (\sqrt{2})T_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Knowing the voltage and ion concentration at various points, it is now possible to construct Table 1 giving the corresponding values of ion current according to eqn. (8).

This Table shows that:

$$\frac{n_{(-T_1)}}{n_{(T_1)}} = \frac{2(T_2 - T_1)}{T_1} = 0{\cdot}828$$

$$\frac{n_{(T_2)}}{n_{(T_1)}} = \frac{T_2 - T_1}{T_1} = 0{\cdot}414$$

and $$\frac{n_{(0)}}{n_{(T_1)}} = 2$$

Such a change in ion concentration is depicted in idealized form in Fig. 4(*b*), the corresponding ramp voltages being shown

Table 1

Time, t	Ramp voltage $V(t)$	Ion concentration $n_{(t)}$	Ion current, i'	Ion current in terms of T_1
$-T_1$	$-\omega V_{max}T_1$	$\frac{2}{3}S(T_2 - T_1)$	$-\frac{2}{3}\omega V_{max}ST_1(T_2 - T_1)C+$	$-\frac{2}{3}\omega V_{max}S0{\cdot}414T_1^2C+$
0	0	$\frac{2}{3}ST_1$	0	0
$+T_1$	$\omega V_{max}T_1$	$\frac{1}{3}ST_1$	$+\frac{1}{3}\omega V_{max}ST_1^2C+$	$+\frac{1}{3}\omega V_{max}ST_1^2C+$
$+T_2$	$\omega V_{max}T_2$	$\frac{1}{3}S(T_2 - T_1)$	$+\frac{1}{3}\omega V_{max}ST_2(T_2 - T_1)C+$	$+\frac{1}{3}\omega V_{max}S0{\cdot}585T_1^2C+$

ı Fig. 4(*a*). The instantaneous currents may be obtained in ke manner:

$$\frac{i'_{(-T_1)}}{i'_{(T_1)}} = -0{\cdot}828 \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$\frac{i'_{(T_2)}}{i'_{(T_1)}} = 0{\cdot}585 \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The current ratios are shown in Fig. 4(*c*) and indicate that a ırrent peak occurs at T_1. Physically, this means that some of ıe ions formed shortly before the voltage reversal are not ɔllected by the time $t = 0$, but reverse their direction of motion ıd are thus collected after 'living' for more than one transit me. Hence a build-up in ion concentration occurs, resulting ı a current peak as the voltage rises. When these transients re over, the voltage across the chamber is sufficiently high to roduce a condition as in a saturated d.c. chamber and the emainder of the current wave is therefore flat, provided that the ɔltage does not rise sufficiently to cause secondary ionization by ɔllision.

(3.3) Effect of Space Charge

In Fig. 4, consider the voltage rise from $V_{(-T1)}$ to zero. Ɔuring this time the positive ions produced drift toward the egative electrode with decreasing velocity, and it is obvious ıat this cloud of ions will eventually be densest near this egative electrode. The converse will hold true if the negative ıarge carriers are also ions, and this form of distribution will ıen produce a field in opposition to the applied field, giving a esultant as shown in Fig. 5(*a*). Thus the ions produced near

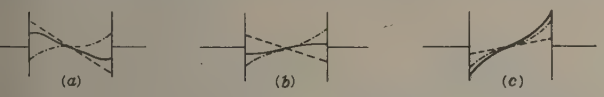


Fig. 5.—Effect of space charge on potential distribution between plates.

– – – – Applied voltage.
—·—·— Space-charge voltage.
——— Resultant voltage.

ıe centre of the active volume will have their velocities much etarded by this effect, and as V approaches zero, the space-ıarge field will become greater than the applied field [Fig. 5(*b*)] ıd a motion of ions in the opposite direction will occur. The et result of this is that the ion concentration at $t = 0$ will be reater than the predicted value and will be markedly hetero-eneous. Furthermore, since the ionic motion due to the space-ıarge field also represents a current in the external circuit, a ırrent waveform which has a leading component may be xpected and therefore $i'_{(0)}$ will not be zero as previously redicted.

A typical current waveshape is shown in Fig. 6, which refers to n RC1 ion chamber as described in Section 4 working in a high eutron flux and having a 70-volt (r.m.s.) 70 c/s supply. The

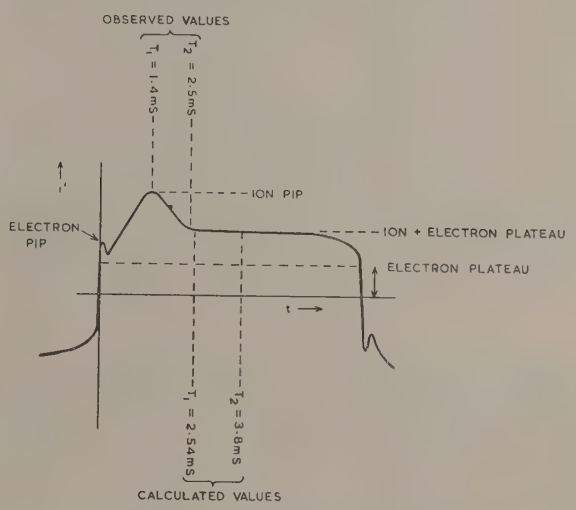


Fig. 6.—Typical current waveform for RC1 chamber under neutron irradiation.

Hydrogen filling at 150 cm Hg.
$V = 70$ volts r.m.s.
$f = 70$ c/s.

hydrogen filling was at a pressure of 150 cm Hg. From this it can be seen that the calculated values of T_1 and T_2 are much larger than the observed values. Such calculated times are, however, maximal, for they refer to ions moving from points very near to the repelling plate. Had the ion distribution been homogeneous, the average transit time would have been given by considering the transit time of an ion formed at the mid-point of the active volume, i.e. at $\frac{1}{2}d$ from either plate. By analogy with eqn. (10) this is

$$\frac{1}{\omega}\arccos\left(1 - \frac{\omega d^2}{4\mu^+ V_{max}}\right)$$

which is 1·24 millisec for the parameters in Fig. 6 ($\mu^+ = 3\,cm^2/volt\text{-}sec$).

Since the batch of ions left uncollected (from the transit time just before the voltage reversal) tends to be concentrated near the new positive-going plate (as explained above), the ion peak should occur later than this, but earlier than T_1. This is shown to be so in Fig. 6, where the observed time at the peak is 1·4 millisec, and the calculated value of T_1 is 2·54 millisec. [The calculated values of T_1 and T_2 were obtained from eqns. (10) and (11).]

When the applied voltage changes direction, the space-charge field will be added to the applied field, as shown in Fig. 5(*c*), and the ions will initially acquire drift velocities in excess of those previously considered. This effect will also help the ion

'pip' to form before T_1, and, in fact, Fig. 6 shows that the initial transient is over at T_1 in this particular case.

The illustration also shows the small electron 'pip' which occurs almost immediately after the voltage reversal. This, as previously stated, is due to the same mechanism as above acting upon the free electrons in the hydrogen, and its early rise and decay is due to their very high mobilities.

The effect of space charge in the saturated a.c.-operated ion chamber is beneficial, for it means that a lower voltage is able to produce a situation analogous to d.c. saturation, but defined in this case as the condition wherein all the ions produced in any one half-cycle are collected before the termination of the next. This is indeed one of the assumptions which must be made if the foregoing treatment is to be valid. It is also a condition which must be met if the response of the chamber is to be linear with flux level, for if a.c. saturation does not occur, a build-up of ions can take place which will be limited only by recombination and diffusion. Since the chamber current is proportional to the ion concentration, it will no longer vary directly with flux level, but will be proportional to the laws governing these two phenomena. A chamber working in the unsaturated mode will have a much higher output current, and its use has been suggested for shut-down channel applications. It is hoped to deal with unsaturated ion-chamber theory and practice elsewhere.

(4) EXPERIMENTAL PROCEDURE

The ion chamber used in the circuit shown in Fig. 2 was the standard A.E.R.E. type RC1. The geometry of this chamber is drawn to scale in Fig. 3, from which it can be seen that the electrodes consist of two concentric cylinders having domed ends. Since the distance between these electrodes (6·2 mm) is small compared with their radii, there will exist a considerable part of the active volume which is subject to a well-defined and (in the absence of space-charge effects) almost uniform field. The plates adjacent to this portion of the volume are coated with boron containing a proportion of the boron 10 isotope. When the chamber is irradiated with neutrons, which ionize only very slightly, owing to their being uncharged, the following transmutation takes place:

$$B^{10} + n^1 \rightarrow Li^7 + He^4$$

The α-particle (helium 4) so produced is strongly ionizing, and generates the charge carriers which are then collected by the applied p.d.

Such a chamber was inserted into an experimental hole in Gleep after the purely capacitive component of current had been balanced out. The variations of response with applied voltage, frequency and flux were plotted and compared with the response under d.c. polarization.

(4.1) Response/Applied-Voltage Characteristics

Since the waveform of the output current was not sinusoidal, the valve millivoltmeter reading, V_0', was in arbitrary units, and the ordinate scale of Fig. 7 is therefore useful only for comparative purposes. The readings, V_0, given by the transfer-function analyser were, however, in r.m.s. millivolts across a load resistance R_F, which in this case was the feedback resistor of the virtual-earth amplifier (A in Fig. 2).

Consider applied voltages of between 10 and 100 volts r.m.s. Fig. 7 shows that, for the lower frequencies, very good plateaux were produced, but that within the applied-voltage limits available, the higher frequencies produced marked peaks. This state of affairs is explicable when it is recalled that the condition for a.c. saturation (as previously defined) depends upon T_1 being less than one half-period, $\frac{1}{2}T_P$. It is apparent that, as the frequency increases or as the peak voltage decreases, T_1 will increase until it exceeds $\frac{1}{2}T_P$, in which case some of the ions produced will never be collected, but will oscillate until removed by diffusion or recombination. These oscillating ions will then be

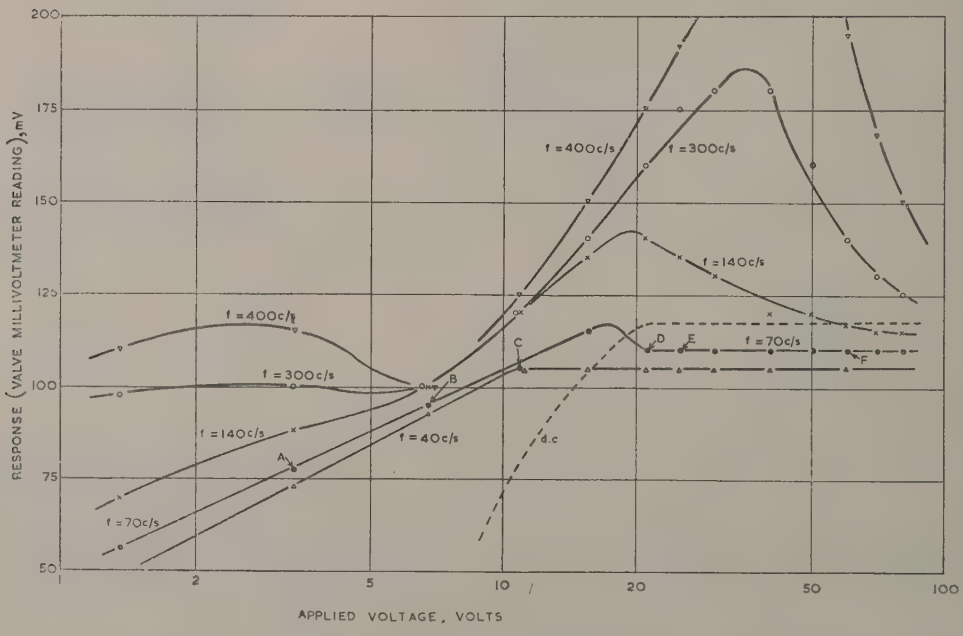


Fig. 7.—Variation of response with applied voltage for RC1 chamber in Gleep.

Hydrogen filling at 150 cm Hg.
$R_F = 1\,M\Omega$
$\phi = 4{\cdot}9 \times 10^6$ n/cm²/sec.

available to produce a current according to eqn. (8), and it is this current which causes the peaks to occur prior to plateau formation in Fig. 7.

Were it not for the effects of space charge, as explained in Section 3.3, the condition for T_1 to be less than $\frac{1}{2}T_P$ could be derived simply from eqn. (10) as follows:

For a.c. saturation to occur,

$$\frac{1}{\omega}\arccos\left(1-\frac{\omega d^2}{\mu^+ V_{max}}\right)<\frac{\pi}{\omega}$$

which gives

$$\frac{\omega d^2}{\mu^+ V_{max}}<2$$

Unfortunately, space-charge effects modify the conditions in the chamber so severely as to make this expression far too inaccurate for practical purposes.

The waveforms in Fig. 8, which correspond to the points

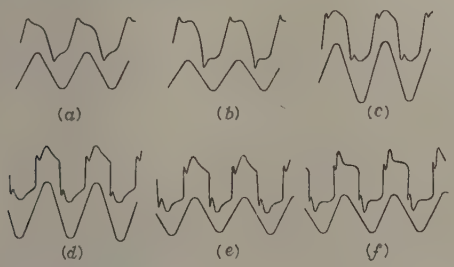


Fig. 8.—Waveforms of chamber current and voltage at 70 c/s corresponding to points A–F in Fig. 7.

(*a*) $V = 3{\cdot}3$ volts.
(*b*) $V = 6{\cdot}7$ volts.
(*c*) $V = 11$ volts.
(*d*) $V = 21$ volts.
(*e*) $V = 25$ volts.
(*f*) $V = 70$ volts.

A–F in Fig. 7, clearly show how the characteristic waveshape is built up as the applied voltage increases at a fixed frequency. They also show how a similar change takes place at voltages below 10 volts r.m.s., and this first 'pip' is due to a mechanism analogous to that described above, but operating upon the free electron cloud. The very high mobilities of electrons make this possible at voltages much lower than those for which the positive-ion phenomena appear.

The curves of the variation of in-phase component of chamber current with applied voltage do not show the low-voltage peaks due to electron collection. These are found to be exhibited only on the curves of the quadrature (leading) current, as would be expected. Otherwise the in-phase curves are of similar form to those in Fig. 7, while the quadrature curves fall to very low values of V_0 as the applied voltage becomes larger.

(4.2) Response/Frequency Characteristics

In Fig. 9 the curve of V_0' corresponds almost exactly with the curve of V_{ip}, the in-phase component of the fundamental. Over the frequency range 40–100 c/s the variation of the common curve shown is only 2%, or 0·033% per cycle.

After the 250 c/s point the curve rises rapidly, again because $T_1 > \frac{1}{2}T_P$. Oscillograms taken at increasing frequencies exhibit the same characteristics as those taken at decreasing voltages, and clearly show that at about 250 c/s the ion 'pip' spreads over the entire half-wave. The ion concentration thus builds up and the chamber current increases proportionally as before.

As this build-up proceeds, the waveform becomes more and more symmetrical, and this condition is reflected in the decreasing

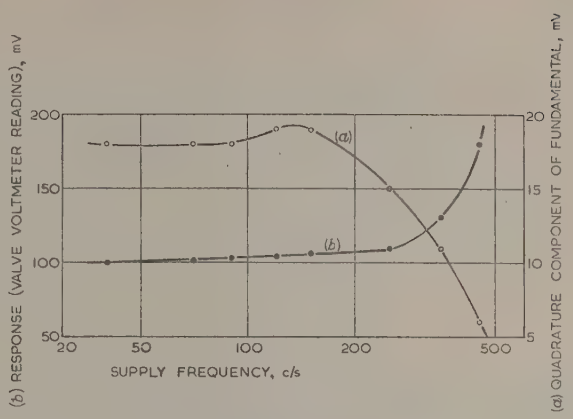


Fig. 9.—Response/frequency curves.

(*a*) Quadrature component.
(*b*) Response.

value of the quadrature component of the fundamental, as seen in Fig. 9. Higher frequencies were not used during this series of tests, since any slight residual capacitance introduced by a small error in the balancing procedure gave rise to a current which at high frequencies was large enough to mask those phenomena due purely to ion motion.

(4.3) Response/Flux-Level Characteristics

Two important effects were observed as the flux increased. First, the minimum voltage for a.c. saturation also increased, as shown in Fig. 10. As in the d.c. case, this effect is due to the

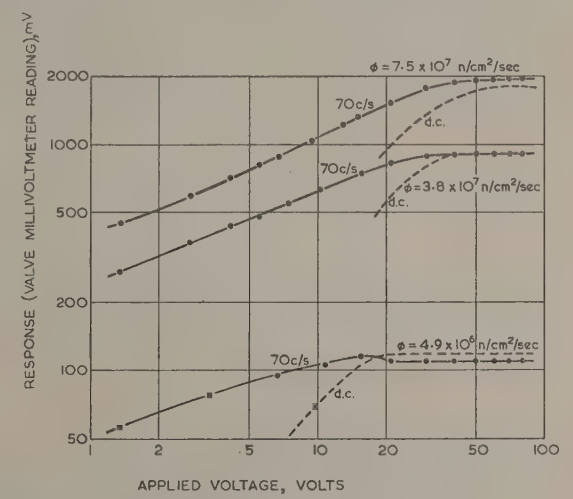


Fig. 10.—Response/applied voltage curves for RC1 chamber in Gleep.
Hydrogen filling at 150 cm Hg.

increased concentration of ions and electrons producing greater recombination and diffusion rates according to eqns. (1) and (3). This means that an increased voltage must be provided so that the ions can be collected before removal by either of these processes. Secondly, since the maximum ion concentration occurs at the top of the 'pip' in the current waveform, this is where the removal of ions by recombination and diffusion will be greatest, and it is for this reason that the observed decrease in the value of r' (pip/plateau ratio) occurred as shown by Fig. 11.

It must here be noted that, since almost the whole of the

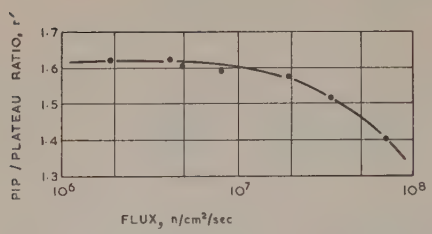


Fig. 11.—Variation of pip/plateau ratio with flux level for RC1 chamber in Gleep.

Hydrogen filling at 150 cm Hg.
$V = 70$ volts.
$f = 70$ c/s.

waveform due to ion collection in a free-electron gas is superimposed upon the plateau due to electron collection, the observed pip/plateau ratio, r', must be less than the true ratio, r.

If the true height of the ion 'pip' is A and the true height of the ion plateau is B, both are superimposed upon an electron plateau of height B, since all electron transients are over and the two types of charge carrier will each contribute half the chamber current in the ion-plateau region. Thus

$$r' = \frac{A + B}{B + B}$$

giving

$$A/B = r = 2r' - 1 \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Fig. 11 shows that $r' = 1{\cdot}61$ before it begins to fall; this gives $r = 2{\cdot}22$.

The excellent linearity of the chamber response with flux level is illustrated in Fig. 12, and it is noteworthy that the reduction in r' does not adversely affect this linearity to any great extent. In the graphs taken using an *RC*1 chamber filled with hydrogen at 15 cm Hg and irradiated in Bepo, complete saturation ceased at about 10^{10} n/cm²/sec, and yet linearity was maintained well above this figure.

(5) EXPERIMENTAL PERFORMANCE WITH AIR FILLING

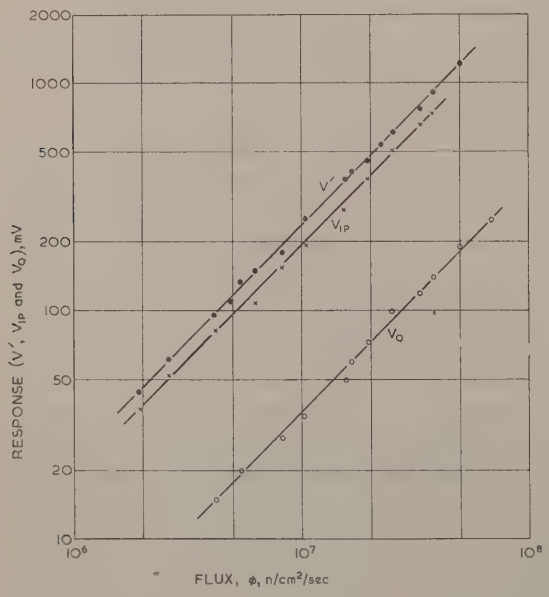
Since no fire hazard is present when using an air filling as opposed to a hydrogen filling, a rubber pipeline was led to the chamber along with the electrical leads, thus making possible a variation in filling pressure with the chamber *in situ*. This system was used only in Gleep while in Bepo a sealed air-filled chamber was used for convenience, owing to differing loading facilities.

(5.1) Response/Filling-Pressure Characteristics

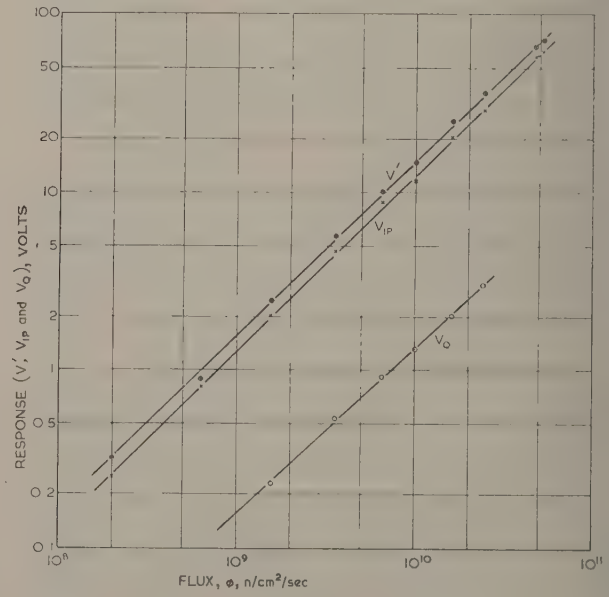
A set of curves taken with the chamber in Gleep is given in Fig. 13, from which it will be seen that the response rises until a maximum is reached at a filling pressure of about 70 cm Hg. The reason for this is that, for each α-particle produced by the $B^{10}(n, \alpha)Li^7$ reaction taking place in the boron lining of the chamber, the collision probability increases with pressure (i.e. molecular concentration). Thus the number of atoms ionized increases until a point is reached where the entire α-particle track is confined within the active volume of the chamber, when all the energy of each particle is used in ionization processes, other than that lost in elastic collisions. Thus a plateau is reached where no more ions can be collected, and this can be clearly seen in the d.c. curve in Fig. 13.

The a.c. curve drops after this point, because, as the ion mobility decreases with increase in pressure, the ratio r also decreases, and since the ion 'pip' on the waveform due to an electro-negative gas is much more pronounced than for a free-electron gas [see Fig. 14(*a*)], the reading of the valve millivoltmeter also drops. (N.B.—The true ratio r will prevail for an electro-negative gas, provided that the charge carriers are of similar mobility and allow both ion peaks to occur concurrently and therefore additively.)

Added to this effect is the increase in loss by recombination of ions at the peak concentrations (engendered by the high concentrations produced at the high pressures involved). It may be that, as the pressure increases still further, a rise in response



(*a*)



(*b*)

Fig. 12.—Response/flux-level characteristic for RC1 chamber.

$V = 70$ volts; $f = 70$ c/s.
(*a*) Chamber filled with hydrogen at 150 cm Hg and irradiated in Gleep.
(*b*) Chamber filled with hydrogen at 15 cm Hg and irradiated in Bepo.

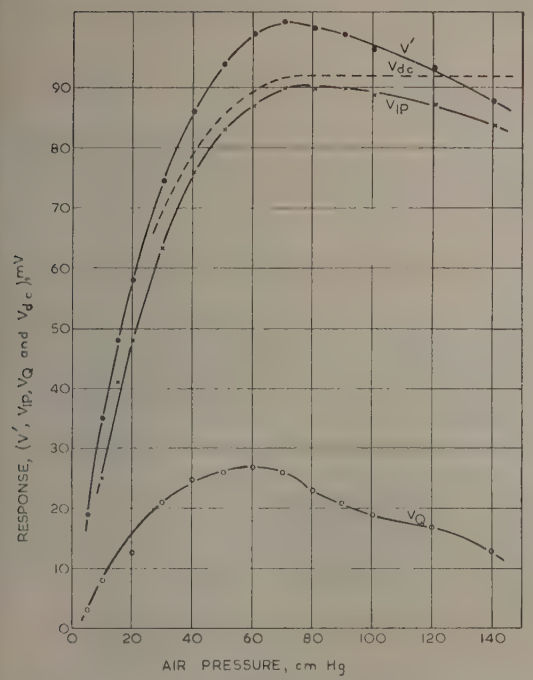


Fig. 13.—Response/pressure curves for RC1 chamber with air filling in Gleep.

might occur, owing to T_1 becoming greater than $\frac{1}{2}T_P$ as the mobility decreases, but this has not been observed as yet, owing to practical limitations.

(5.2) Output Waveforms

As expected, the waveform of the chamber current showed no electron 'pip' [see Fig. 14(*a*)]. This was due to the high electron

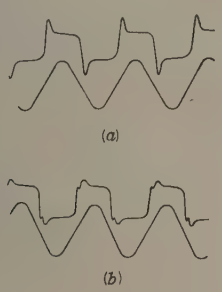


Fig. 14.—Current waveforms from RC1 chamber with air filling at 15 cm Hg in Bepo.

(*a*) First day.
(*b*) Fifteenth day.

attachment coefficient of the oxygen in the air, and since positive and negative ions in air have very similar mobilities, the ion 'pip' was found to be much higher relative to the plateau than the case for the hydrogen filling. During this phase in the experiments it was observed that the current waveform slowly changed character when the chamber was left inside the pile for some days in a flux of the order of 10^{10} n/cm^2/sec. A tracing of the final waveform exhibited by the chamber which produced the trace of Fig. 14(*a*) is shown in Fig. 14(*b*).

The presence of an electron 'pip' will be seen, as will the decrease in height of the ion 'pip'.

Investigations[6] have recently been made into this phenomenon at the Atomic Energy Research Establishment and it is now thought to be due to depletion in the oxygen content of the air by the formation of ozone and oxides of nitrogen. It has been shown that a free-electron type of waveform can occur with a mixture consisting of nitrogen containing up to 10% of oxygen, which accounts for the apparent anomaly of the impossibility of total oxygen clean-up by the small amount of boron present. It has also been demonstrated that only some 0·5% of oxygen in a hydrogen chamber is sufficient to produce a single-'pip' waveform.

(6) CONCLUSION

It will be appreciated that the work described is, in essence, an early attempt to define the conditions under which a typical ion chamber can be usefully operated when supplied with an alternating voltage. This means that much further work remains to be done, but it is already apparent that such a chamber is fully capable of being operated in a linear mode with respect to flux level and, this being so, could be used in the high-flux regions of nuclear reactors as a measurement or control transducer.

The range over which the output would remain linear is dependent upon the voltage and frequency of the supply, and, for the best saturation characteristics, a filling of a light free-electron gas such as hydrogen is obviously preferred. Such a filling also gives a low value of r', which makes for simplicity in the detector.

The speed of response of the detector system was determined empirically in this case, for in order to maintain stability in the virtual-earth amplifier with the large capacitance of 400 pF appearing at the grid, it was necessary to shunt the 1-megohm feedback resistor with a small capacitor which restricted the frequency response. Under these conditions, the system was found to be 3 dB down at 4·5 kc/s, which corresponded to a rise time of better than 1 millisec. This was entirely satisfactory and quite comparable with the performance of an equivalent d.c. system.

For use with lower fluxes, more sophisticated backing-off methods are obviously necessary, and in this connection, a γ-ray compensated chamber might well be adapted to provide capacitive current compensation also.

It is not only in the saturated mode that the a.c. ion chamber will usefully operate, for when the ion concentration rises, owing to lack of collection, a greatly increased current can be obtained. It is hoped to deal with the basics of this mode of operation at a later date.

(7) ACKNOWLEDGMENTS

The author is greatly indebted to Dr. J. H. Mitchell and Mr. J. R. Adams of Ericsson Telephones, Ltd., to Messrs. R. J. Cox, W. Abson and D. Harrison of A.E.R.E. Harwell; and to Professor J. E. Parton of the University of Nottingham, who in various ways made this research project possible. The work was carried out under a research contract placed by A.E.R.E. Harwell, with Ericsson Telephones Ltd. and the University of Nottingham.

(8) REFERENCES

(1) KIEHN, N. M.: 'A.C. Operation of Ionization Chambers' (Argonne National Laboratory, U.S.A.), A.E.C.U. Report 1630.
(2) WARD, A. H.: 'Low-Frequency Operation of Balanced Ion Chambers in an A.C. Bridge', A.E.R.E. Report EL/R1107.
(3) COBINE, J. D.: 'Gaseous Conductors' (McGraw-Hill, 1941), p. 98.

(4) VON ENGEL, A.: 'Ionised Gases' (Clarendon Press, 1955).
(5) COBINE, J. D.: 'Gaseous Conductors' (McGraw-Hill, 1941), p. 105.
(6) ROBERTS, D. L.: A.E.R.E. Research Report R2905, April, 1959.

(9) APPENDIX: DERIVATION OF COLLECTED FRACTIONS

The treatment given in Section 3.2 is very approximate and serves only to indicate the nature of the phenomena occurring in the chamber. Certain points other than the deviations due to space-charge effects can, however, be elucidated; one is the contention that, as the voltage falls to zero, one-third of a group of ions formed over a transit time are collected, while as the voltage rises from zero, two-thirds of such a group are collected.

Consider a transit time, T_2–T_1, which occurs during a positive half-cycle of voltage. During a small interval dt at t (in this transit time) a group of Sdt ions will be generated. Now, at some point x in the space between the plates, an ion will be formed which is just collected at T_2; therefore, all ions formed behind a plane passing through the ion and lying parallel to the plates will fail to be collected by the time T_2. Therefore, of the group of ions formed, $(x/d)Sdt$ will fail to be collected by the time T_2, and the total number of ions formed between T_1 and T_2 not collected is

$$\int_{T_1}^{T_2} \frac{xS}{d} dt \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

But $$x = \frac{\mu V_{max}}{d} \int \sin \omega t dt = -\frac{\mu V_{max}}{\omega d} \cos \omega t + C$$

When $t = T_1$, $x = 0$, giving $C = \frac{\mu V_{max}}{\omega d} \cos \omega T_1$

Therefore $$x = \frac{\mu V_{max}}{\omega d} (\cos \omega T_1 - \cos \omega t)$$

Thus the total not collected is

$$\frac{\mu V_{max} S}{\omega d^2} \int_{T_1}^{T_2} (\cos \omega T_1 - \cos \omega t) dt$$

$$= \frac{\mu V_{max} S}{\omega d^2} \left[(T_2 - T_1) \cos \omega T_1 - \frac{1}{\omega} (\sin \omega T_2 - \sin \omega T_1) \right]$$

We may now substitute

$$\frac{\mu V_{max}}{\omega d^2} = \frac{1}{\cos \omega T_1 - \cos \omega T_2}$$

(obtained by putting $x = d$ when $t = T_2$).

The fraction not collected is now found by dividing the resulting expression by the total generated, $S(T_2 - T_1)$, giving the fraction not collected as

$$\frac{(T_2 - T_1) \cos \omega T_1 - \frac{1}{\omega} (\sin \omega T_2 - \sin \omega T_1)}{(T_2 - T_1)(\cos \omega T_1 - \cos \omega T_2)} \quad . \quad (21)$$

The situation at the origin can be determined by letting $T_1 = 0$ in eqn. (21), which then gives the fraction not collected as

$$\frac{1 - \frac{1}{\omega T_2} \sin \omega T_2}{1 - \cos \omega T_2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

If T_2 is very small, this expression tends to one-third, thus showing that two-thirds of the group of ST_2 ions are collected.

If now a similar analysis is made for a ramp function, $V = mt$ or $V = -mt$, it is found that exactly two-thirds of the ions produced in any transit time are collected for the rising-voltage case, and exactly one-third are collected for the falling-voltage case. In other words, the portion of sinusoidal voltage wave rising from $t = 0$ gives results similar to those of a rising ramp function. When the applied voltage reaches its peak at $t = \pi/2\omega$, the fraction collected over a transit time whose mid-point lies at $t = \pi/2\omega$ can be determined by making the following substitution in eqn. (21):

Let $$T_1 = \left(\frac{\pi}{2\omega} - \delta T\right) \text{ and } T_2 = \left(\frac{\pi}{2\omega} + \delta T\right)$$

where the transit time $T_2 - T_1 = 2\delta T$.

Using these substitutions, eqn. (21) gives one-half as the fraction not collected, which corresponds to the case where $T_1 \rightarrow T_2$, and the top of the voltage wave may be approximated to a straight horizontal line.

Finally, when the voltage wave falls to zero, it is convenient to consider the transit time immediately preceding $t = 0$ and to substitute $-T_1$ and $T_2 = 0$ into eqn. (21). This gives the fraction not collected as

$$\frac{T_1 \cos \omega T_1 - \frac{1}{\omega} \sin \omega T_1}{T_1 (\cos \omega T_1 - 1)}$$

Therefore the fraction collected may be found by subtracting this from unity, giving

$$\frac{1 - \frac{1}{\omega T_1} \sin \omega T_1}{1 - \cos \omega T_1} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

This expression tends to one-third when T_1 is very small, giving two-thirds as the fraction not collected. This corresponds to the case for a ramp function of voltage $V = -mt$, as expected.

DISCUSSION ON 'A NOVEL, HIGH-ACCURACY CIRCUIT FOR THE MEASUREMENT OF IMPEDANCE IN THE A.F., R.F. AND V.H.F. RANGES'*

Mr. W. H. P. Leslie (*communicated*): The author, in replying to points raised in the discussion, would have done better to admit an amount of error rather than to assert that all was well and justified by the mathematics.

He bases the mathematics on his Fig. 1, the basic circuit with two mutual inductors. Although his mathematics (after the adjustments detailed in the discussion) show that his bridge will work, a few minutes' thought will show that it will not. This is more obvious if the circuit is redrawn as in Fig. A.

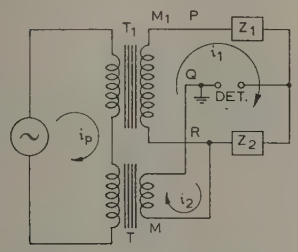


Fig. A.—Two transformers.
$i_2 = 0$ if balanced.

The bridge, if it will balance, depends on an induced e.m.f. due to M which will provide an effective tap on the induced e.m.f. due to M_1. It is shown by the author that if $Z_1 = Z_2$ then $M_1 = 2M$. Now, it is seen that the upper transformer, T_1, is connected to a load $Z_1 + Z_2$, whereas the lower transformer T, is open-circuited at the proposed balance. The primary current then flows in a reflected load in the upper primary and in a high impedance (shunt inductance) in the lower primary. Under these conditions the e.m.f. in the windings of T would be very much higher than that in the windings of T_1, so that the potential at Q would lie well above that at P (relative to R). Balance could only occur with Z_1 a negative impedance.

The author's mathematics tell a different story—he has omitted a term $+j\omega L_1 I_1$ from the right-hand side of his first equation (1) (L_1 is the shunt inductance of the secondary), or alternatively his Z_4 contains such a term and thus his $Z_4 > Z_1$ and so his eqn. (2),

$$\frac{M - M_1}{M} Z_2 = Z_1 + Z_4$$

is not very helpful unless he measures the dominant Z_4 term.

He then goes on to circuits based on his Fig. 2, on which all his tests were made. This circuit can be redrawn as in Fig. B, which shows that, whilst this is indeed a novel arrangement, it is a very close relative of the simple ratio-arm bridge normally shown as in Fig. C.

The mathematical analysis is now reasonably correct, apparently by accident, being based by analogy on the erroneous analysis of Fig. 1.

His bridge does not reject the effect of stray shunt capacitance

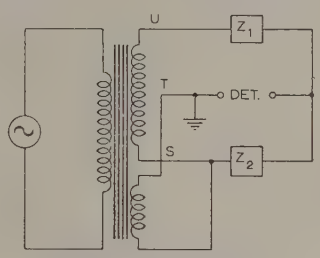


Fig. B.—One transformer.

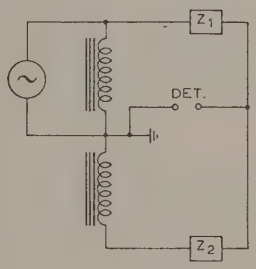


Fig. C.—One transformer (Lynch).

of the unknown so effectively as does the bridge described by Lynch, in which the shunt susceptance is supplied directly by the oscillator, the transformer being reserved for the production of a proportional voltage of opposite phase for the standard (Z_2) (see Fig. C).

The author, in referring to his Fig. 3, states that capacitors C_1 and C_4 (between U and T in Fig. B) shunt impedance Z_1, whilst C_2 and C_3 (between S and T) do not shunt Z_2. Since C_1, C_4 and Z_1 are arranged in an identical fashion (with respect to earth) to C_2, C_3 and Z_2, his statement cannot be accepted. His bridges, as exemplified in Fig. B, all suffer from the winding SU carrying the load current and thus having an e.m.f. slightly greater than the terminal voltage, whilst winding ST has no such voltage drop unless the stray shunt capacitance happens to approach the value of the unknown.

To summarize: if the author wishes to use mathematics to substantiate his claims he will require to take account of more factors than appear in his paper. His proposed winding arrangement seems inferior to the tapped transformer: his windings will certainly have a higher resistance for a given space, as roughly 50% more turns are needed.

Dr. D. Karo (*in reply*): There is no error, not even part of an error, to admit to in the paper as modified in my reply to the first discussion.*

I am shocked by Mr. Leslie's contradiction of a correct mathematical development based on sound assumptions and backed by experiment. One cannot treat the circuit of Fig. 1,†

* Karo, D.: Paper No. 2709 R, November, 1958 (see 1958, 105 B, p. 505, and 1959, 106 B, p. 435).

* 1959, 106 B, p. 435.

† 1958, 105 B, p. 506.

with two mutual inductors in series, as tapped chokes in series or as a potentiometer arrangement. Bringing into the relevant equations the reflected impedances from primary to secondary, or vice versa, will not make any difference.

M_1 and M are mutual inductances, i.e. flux linkages from the respective primaries into the respective secondaries per unit current in the two primaries (which are connected in series). When M_1 and M are multiplied by ωI_p the terms represent the e.m.f.'s induced in the secondaries.

M_1 and M are constant and the reflected impedances do not matter.

Let the impedance of the primary of M_1 (Fig. 1) be Z and the p.d. across it V_1.

We can write

$$V_1 = ZI_p + jM_1\omega I_1; \quad -jM_1\omega I_p = (Z_1 + Z_2 + Z_4)I_1$$

Hence,

$$I_1 = \frac{-jM_1\omega I_p}{Z_1 + Z_2 + Z_4}; \; I_p = \frac{V_1}{Z} - \frac{M_1^2\omega^2 I_p}{Z(Z_1 + Z_2 + Z_4)}$$

$$\frac{V_1}{I_p} = Z + \frac{M_1^2\omega^2}{Z_1 + Z_2 + Z_4}$$

The last expression contains the impedance Z plus the impedance reflected from secondary to primary.

Eqn. (1) can therefore be written

$$(Z_1 + Z_2 + Z_4)I_1 = -jM_1\omega\left[\frac{V_1}{Z} - \frac{M_1^2\omega^2 I_p}{Z(Z_1 + Z_2 + Z_4)}\right]$$

$$-Z_2 I_1 = jM\omega\left[\frac{V_1}{Z} - \frac{M_1^2\omega^2 I_p}{Z(Z_1 + Z_2 + Z_4)}\right]$$

Hence

$$\frac{Z_1 + Z_2 + Z_4}{-Z_2} = \frac{M_1}{M}$$

as in the paper.

Looking from the secondary of M_1 one obtains

$$I_1 = \frac{-j(V_1 - ZI_p)}{M_1\omega}$$

Hence eqn. (1) becomes

$$-jM_1\omega I_p = (Z_1 + Z_2 + Z_4)\left[\frac{-j(V_1 - ZI_p)}{M_1\omega}\right]$$

$$jM\omega I_p = -Z_2\left[\frac{-j(V_1 - ZI_p}{M_1\omega}\right]$$

Therefore, again,

$$\frac{M_1}{M} = \frac{Z_1 + Z_2 + Z_4}{-Z_2}$$

The conclusions reached in the paper are therefore correct: the e.m.f. in the lower secondary will not be higher than in the upper one, Z_1 will not have to be a negative impedance, the term $j\omega L_1 I_1$ (in Mr. Leslie's notation) has not been omitted, and Z_4 need not contain such a term and is not the dominant one. Z_4, as the balance equations show, is known. Incidentally, I have shown that when $M_1 = 2M$ then $Z + Z_4 = Z_2$, and not, as Mr. Leslie says, $Z_1 = Z_2$.

I am glad that Mr. Leslie agrees that my circuit of Fig. 2 is novel; it may appear to be a very close relative of the one shown in his Fig. C, but only in his drawing, which is incomplete in many respects.

Mr. Leslie then goes on to state that the mathematical treatment is reasonably correct. This will not do: either the mathematics are correct for Figs. 1 and 2 (as they are indeed) or they are not.

The statement that my circuit does not reject the stray shunt capacitance of the unknown as well as Lynch's circuit is quite without foundation, as is clear from the paper and the discussion; in any case I did not make any statement to this effect.

I do not state, as Mr. Leslie alleges, that the capacitances C_2 and C_3 do not shunt Z_2 (Figs. 3 and 4). These capacitances are dealt with in a different way from C_1 and C_4, the reasons having been given in the paper and discussion.

The circuits described do not suffer from the upper winding carrying the load current or from anything else that Mr. Leslie imagines.

As regards the last paragraph in his communication, all the relevant factors have been taken account of. The statement that the winding arrangement is inferior to the tapped transformer that it will have a higher resistance and that 50% more turns are needed for a given space is contrary to the facts.

Finally, the tests were done on the circuits of Figs. 1 and 2 with the same very satisfactory results.

MONOGRAPHS PUBLISHED INDIVIDUALLY

Summaries are given below of monographs which have been published individually, price 2s. (post free). Applications, quoting the serial numbers as well as the authors' names, and accompanied by a remittance, should be addressed to the Secretary. For convenience, books of five vouchers, price 10s., can be supplied.

Voltage and Current Transformation Matrices. Monograph No. 355 E.

I. CEDERBAUM, Ph.D.

Transformation matrices relating two adequate systems of simple network co-ordinates such as node-pair voltages or link currents belong to the class of unimodular or E-matrices. The paper makes a distinction between the matrices corresponding to such voltage and current transformations, and different sets of necessary conditions are derived for each type.

Since the loop- and cut-set-to-branch incidence matrices are closely related to transformation matrices, the discussion proposes new sets of necessary conditions for incidence matrices corresponding to systems of node-pair voltages or link currents. Examples are given of matrices which whilst representing a voltage transformation cannot represent a current transformation and vice versa. Another example shows an E-matrix which can represent neither a voltage nor a current transformation.

The Stabilization of Control Systems with Backlash using a High-Frequency On–Off Loop. Monograph No. 356 M.

E. A. FREEMAN, B.Sc., Ph.D.

The paper introduces a method of stabilizing control systems which have backlash in their control sequence. The technique described employs an auxiliary loop to drive the motor across the backlash whenever the motor and load tend to separate. Requirements of the auxiliary loop are deduced and a phase-plane analysis is developed for a second-order position-control system with this loop operative. Analysis shows that the system with backlash is effectively linearized. It is also shown that impacts between motor and load are avoided.

To establish the feasibility of the method an analogue computer study is presented. Results from the analogue show that the technique leads to considerable improvement in the step-function response. Variation of the system damping and of the ratio of inertias is also investigated. It is found that the system stabilized by the technique behaves essentially as a linear system.

The Conductivity of Oxide Cathodes: Part 8. Current-Dependent Matrix Dissociation. Monograph No. 357 E.

G. H. METSON, M.C., D.Sc., Ph.D., M.Sc., B.Sc.(Eng.), and EDITH MACARTNEY, M.Sc., B.Sc.

In the present Part an attempt is made to determine the nature of the dissociative action which accompanies the passage of a current through a barium-strontium-oxide matrix at 1 020° K. Two identical oxide systems operating at a common temperature and passing the same quantity of electricity but at different rates might be expected to suffer the same mass of oxide dissociation. Experiment shows, however, that such is not necessarily the case and that dissociation mass is determined by rate of application of electricity rather than by the total quantity of electricity.

The products of dissociation are shown to be in ionic form, and experimental arrangements are made for the separate collection of the ions in chemical form on a relatively massive scale.

A working hypothesis is offered in explanation of the experimental observations.

Radiation from Discontinuities in Strip-Line. Monograph No. 358 E.

L. LEWIN.

A method of calculation, which should be adequate for strip-lines of small spacing, is proposed for the evaluation of the radiation of line-above-ground configurations. It is shown that an open-circuit is appreciably worse, from the point of view of radiation loss, than a short-circuit with a right-angle corner at an intermediate level. The radiation from a matched post is slightly worse than from the short-circuit, but it should be capable of considerable improvement by the use of a terminating frame antenna. Formulae are also given for the effects of reactive posts, with applications to a simple resonator.

Some Observations on Waveguide Coupling through Medium-Sized Slots. Monograph No. 359 E.

L. LEWIN.

A quasi-static transmission-line method, as used in antenna theory, and modified to take into account effects of radiation damping, is applied to the problem of waveguide coupling through slots which may be large enough to exhibit resonance effects. It is shown that the waveguide environment significantly alters the value to be used for the slot characteristic impedance, although this quantity cannot itself be calculated by the present analysis. The method is applied in detail to the case of axially-coupled waveguides and to guides coupled, via a common broad wall, through crossed slots. It is shown that a narrow slot needs to be very near resonance to give a coupling as large as that from a circular hole of comparable dimensions.

A New Design Method for Phase-Corrected Reflectors at Microwave Frequencies. Monograph No. 360 E.

S. CORNBLEET, B.Sc.

A method is given for the design of a class of wide-angle reflectors in which the aberrations are reduced by coating the reflector surface with a dielectric. At microwave frequencies this dielectric may take the form of an array of metal plates or waveguides and the constraint imposed by this form greatly simplifies the analysis. The path of the feed point is chosen so that only selected rays from it are equally phased. It is then found that both of the reflector's profiles and the path of the feed point can be described by a single parameter. The residual aberrations of several cylindrical systems, including the corrected parabola and circle, are analysed to find a reflector for which the scanning arc is the circle with centre at the vertex of the reflector. A refocusing procedure, which is found to be necessary, produces this scanning arc. Experimental results in agreement with the theory are given.

It is shown that the design principle can be used to programme a step-wise procedure for the design of corrected surfaces using non-constrained natural dielectric coatings.

The Effect of an Additional Non-Linearity on the Performance of Torque-Limited Control Systems subjected to Random Inputs. Monograph No. 361 M.

J. L. DOUCE, M.Sc., Ph.D., and R. E. KING, M.Sc.

The paper discusses a technique for improving the response of a saturating servo mechanism subjected to random signals. It is shown that a non-linear error detector gives a considerable reduction in error magnitude for a large range of input signals. It is possible to design the additional non-linearity to optimize the performance of the system for all input magnitudes. This form of non-linearity has previously been shown to effect improvement in the response of such systems to step-function and sinusoidal inputs.

Experimental work verifies the predicted results for a particular system, and shows that considerable latitude is permitted for the characteristic of the additional non-linearity.

Power Flow and Negative Wave Impedance in the Dielectric-Rod Waveguide. Monograph No. 362 E.

E. F. F. GILLESPIE, M.Eng.

In most waveguides of practical interest the wave impedance in the direction of propagation of the wave is positive and its magnitude does not vary with position. It is shown that, in general, a guiding structure supporting a hybrid mode, i.e. a mode having axial components of both the electric and magnetic fields, can present a

negative wave impedance. The case of the EH_{11} mode on a dielectric rod is then considered, and it is shown that for this mode the wave impedance is negative over part of the transverse plane and hence leads to regions of negative power flow if the usual interpretation of the Poynting vector is employed. The total power flow, however, is still positive, since the negative power density and the negative wave impedance occur only in a restricted part of the cross-section.

The orthogonality of the fields is also discussed, and it is shown that for hybrid modes the transverse electric and magnetic fields are not at right angles to each other.

The Surface-Wave Aerial. Monograph No. 363 E.

W. Hersch, Ph.D., B.Sc.(Eng.).

The radiation from the open-circuited end of an externally dielectric-coated metallic waveguide can be controlled by varying the size of the guide, the thickness and/or the dielectric constant of the coating. A new type of aerial designed around this principle is given the name 'surface-wave aerial' and radiation-pattern measurements are used to confirm the theory underlying this type of radiator.

According to its mode of operation it belongs to the category of end-fire aerials, which are briefly reviewed to show that surface-wave aerials occupy a place in their own right amongst the many possible arrangements that utilize the end-fire effect to produce a directional radiation pattern.

The theory of the surface-wave aerial is developed in detail, a necessary preliminary step being a full theoretical analysis of the properties of the first-order cylindrical surface wave. It is shown that a dielectric-coated cylinder which is approximately a wavelength in circumference can act as a waveguide for higher-order surface waves, of which the first order is an example.

The 'characteristic equation' is determined for the general case from which the cut-off frequency, propagation coefficient and conditions under which propagation can take place are derived in turn.

Two specific cases are evaluated numerically and the results are used to calculate the polar diagrams of surface-wave aerials operating at 9 Gc/s.

The wavelength constant as well as the continuous radiation loss of a surface waveguide for which $\lambda_g/\lambda_0 \simeq 1$ are measured directly and the results obtained are used to account for the radiation pattern of very long aerials. In conclusion, an outline of future work is given.

The Characteristics of the Trigatron Spark-Gap at Very High Voltages. Monograph No. 364 M.

T. E. Broadbent, M.Sc., Ph.D.

Curves are given showing the working range and time-lag characteristics of a trigatron spark-gap working in air at voltages up to 1 MV. It is shown that, for voltages of this magnitude, a single-stage trigatron spark-gap of suitable design provides a simple and convenient method of chopping the voltage at any required instant. Factors which affect the performance of the gap are discussed.

An Experimental Study of the Slot Aerial and the Three-Element Colinear Array of Slot Aerials. Monograph No. 365 E.

Ronald King, Ph.D., and Gilbert H. Owyang, Ph.D.

The principle of complementarity is reviewed with particular reference to its application to the study of slot transmission lines and slot aerials. The conventional techniques for making measurements on transmission lines are shown to be applicable to a two-slot line that is used to centre-drive a slot aerial. The complementary normalized impedance of such an aerial as a function of its half length, and the distributions of the electric and magnetic field along it, have been measured and are described. The experimental results for the slot aerial are compared with the approximately complementary ones derived theoretically for a centre-driven cylindrical aerial. The effect of the finite thickness of the ground screen in which the slot line and aerial are cut is discussed. A similar study is described for a three-element colinear array of slot aerials.

A Theory of Steady Forces in Variable-Parameter Networks. Monograph No. 366 M.

W. E. Smith, B.Sc.

Passive linear reciprocal networks with elements dependent upon a generalized co-ordinate x are investigated in order to evaluate the steady component of the corresponding generalized force $(F_x)_{av}$. Methods are found for expressing this force in terms of parameters measurable at the input terminals of the network.

The analysis is particularly relevant to the absolute calibration of a class of square-law electrical measuring instruments, but the formalism is also applicable to other network systems where impedance is a useful concept.

Further Theory of a Certain Continued Fraction. Monograph No. 367 M.

O. P. D. Cutteridge, M.Sc.(Eng.), Ph.D.

The paper develops further theory of a certain type of continued fraction relevant to the problem of determining the character of the zeros of a polynomial. Two theorems provide tests for the number of positive zeros, real zeros and pairs of conjugate complex zeros of a real polynomial. Two numerical examples are included, one of which shows the application of the method to a problem in linear-network theory.

Orthogonal Codes. Monograph No. 369 E.

H. F. Harmuth, Dipl.Ing., Dr.Tech.

Code alphabets whose characters can be represented by a finite sequence of digits of value $+1$ or -1 have been extensively investigated. The characters of these binary codes may be considered to be superpositions of orthogonal functions, one for each digit, which are multiplied by $+1$ or -1. A character of an orthogonal code consists of one function of a set of orthogonal functions multiplied by $+1$ or -1.

AN ANNOUNCEMENT FROM ULTRA . . .

Since the Special Products Division of Ultra Electric Limited was formed in 1946, the demand for its specialised services in the field of electronic engineering has grown steadily. To meet the increase in the number and complexity of these commitments, a programme for expansion and re-organisation was recently completed and the Division has been re-formed as an autonomous Company known as ULTRA ELECTRONICS LIMITED.

The Company has first-class facilities for research, development and production. Current activities include work on radio and radar system applications, magnetic amplifier servo controls, aircraft intercommunication, submarine detection, and computer peripheral equipment.

The Continental Connector Division of the new Company manufactures the precision miniature multi-way connector of that name.

ULTRA

ULTRA ELECTRONICS LIMITED
WESTERN AVENUE · LONDON · W3 · Tel: ACOrn 3434

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. ELECTRONIC AND COMMUNICATION ENGINEERING (INCLUDING RADIO ENGINEERING), MARCH 1960

CONTENTS

	PAGE
Some Comments on the Classification of Waveguide Modes A. E. KARBOWIAK, B.Sc.(Eng.), Ph.D.	85
An Analogue Electronic Multiplier using Transistors as Square-Wave Modulators P. GLEGHORN, B.Sc.(Eng.)	94
Theory of the Travelling-Wave Parametric Amplifier Prof. A. L. CULLEN, Ph.D.	101
Some Properties of Travelling-Wave Resonance J. R. G. TWISLETON, B.Sc.	108
Saturation Effects in a Travelling-Wave Parametric Amplifier A. JURKUS, B.Sc.A., and P. N. ROBSON, B.A.	119
Discussion on the above three papers and on 'The Gain of Travelling-Wave Ferromagnetic Amplifiers'	123
The Propagation of High-Frequency Radio Waves to Long Distances F. KIFT	127
V.H.F. Field-Strength Measurements over Paths in the Irish Sea involving Mountain Obstacles J. K. S. JOWETT, B.Sc.(Eng.)	141
A Method of Amplitude and Phase Measurement in the V.H.F.–U.H.F. Band. G. D. MONTEATH, B.Sc., D. J. WHYTHE, B.Sc.(Eng.), and K. W. T. HUGHES, B.Sc.	150
Fundamental and Harmonic Distortion of Waves Frequency-Modulated with a Single Tone R. G. MEDHURST, B.Sc.	155
Post Office Valves for Deep-Water Submarine Telephone Repeaters M. F. HOLMES, B.Sc., and F. H. REYNOLDS, B.Sc.(Eng.)	165
High-Power Transmitting Valves with Thoriated Filaments for Use in Broadcasting. H. S. WALKER, M.B.E., W. H. ALDOUS, B.Sc., R. G. ROACH, B.Sc., J. B. WEBB, D.F.H., and F. D. GOODCHILD, B.Sc.(Eng.)	172
An Improvement to the Electron-Trajectory Tracer J. VINE, M.Sc., and R. T. TAYLOR	181
Improvements in the Precision Measurement of Capacitance G. H. RAYNER, B.A., and L. H. FORD, M.Sc.	185
Ferroresonant Systems of Circuit Logic Prof. J. G. SANTESMASES, M. ALIQUE, Dr.Sc., and J. L. LLORET, Bs.Sc.	190
Impedance/Frequency Characteristics of Glow-Discharge Reference Tubes F. A. BENSON, D.Eng., Ph.D., and P. M. CHALMERS, B.Eng.	199
The Principle of Reversed Lag applied to On-Off Temperature Control H. SUTCLIFFE, M.A.	209
The A.C. Operation of Ion Chambers at High Fluxes J. WATSON, B.Sc., S.M., Ph.D.	216
Discussion on 'A Novel, High-Accuracy Circuit for the Measurement of Impedance in the A.F., R.F. and V.H.F. Ranges'	225
Monographs published individually	227

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 4001 R, December, 1954 (102 B, p. 1234).

★ There are many members and former members of The Institution who are finding life difficult. **When remitting your membership subscription,** please help them by sending a donation, or annual subscription preferably under deed of covenant, to

THE BENEVOLENT FUND

The object of the Fund is to afford assistance to necessitous members and former members (of any class) of The Institution of Electrical Engineers who have paid their subscriptions for at least five years consecutively or compounded therefor, and to the dependants of such members or former members.

Subscriptions and Donations may be sent by post to

THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF ELECTRICAL ENGINEERS, SAVOY PLACE, LONDON, W.C.2

or may be handed to one of the Local Hon. Treasurers of the Fund.

★ *Though your gift may be small, please do not hesitate to send it*

Local Hon. Treasurers of the Fund:

EAST MIDLAND CENTRE *H. J. McLean*	SCOTTISH CENTRE *R. H. Dean, B.Sc.Tech.*
IRISH BRANCH *A. Harkin, M.E.*	NORTH SCOTLAND SUB-CENTRE. . . . *P. Philip*
MERSEY AND NORTH WALES CENTRE *D. A. Picken*	SOUTH MIDLAND CENTRE . . . *H. M. Fricke*
TEES-SIDE SUB-CENTRE *W. K. Harrison*	RUGBY SUB-CENTRE *P. G. Ross, B.Sc.*
NORTH-EASTERN CENTRE . . . *J. F. Skipsey, B.Sc.*	SOUTHERN CENTRE *J. E. Brunnen*
NORTH MIDLAND CENTRE *E. C. Walton, Ph.D., B.Eng.*	WESTERN CENTRE (BRISTOL) . . . *A. H. McQueen*
SHEFFIELD SUB-CENTRE *F. Seddon*	WESTERN CENTRE (CARDIFF) . . . *E. W. S. Watt*
NORTH-WESTERN CENTRE . *E. G. Taylor, B.Sc.(Eng.)*	WEST WALES (SWANSEA) SUB-CENTRE *O. J. Mayo*
NORTH LANCASHIRE SUB-CENTRE . *H. Charnley*	SOUTH WESTERN SUB-CENTRE . . *W. E. Johnson*
NORTHERN IRELAND CENTRE . . *G. H. Moir, J.P.*	

Members are asked to bring to the notice of the Court of Governors any deserving cases of which they may have knowledge.

THE BENEVOLENT FUND

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* COVent Garden 1871. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.